Position: Multi-Agent AI for Emergent Behavior Requires Mean-Field Learning

Anonymous Author(s)
Affiliation
Address
email

Abstract

As AI agents scale to large populations, traditional coordination mechanisms face intractable complexity. In multi-agent settings, *agentic emergence* is a term commonly used to characterize unpredictable collective behaviors at scales of multi-agent interactions, but is not well-defined. **This position paper argues that grounding agentic emergence in Artificial Social Intelligence and Mean-Field Theory is necessary for transforming illusory heuristics to rigorous scaling laws**. Current multi-agent interaction paradigms face exponential sample complexity barriers that fundamentally limit scalability due to combinatorial interaction spaces. We argue that mean-field learning (MFL) is not merely an optimization technique but a principled and increasingly necessary requirement for reliable large-population orchestration. We support this position by pointing to recent advances in sample complexity driven by game theory and MFL with example illustrations in agentic coding and robotics. Finally, we offer a rationalized definition of emergence and outline a future research agenda where mean-field learning turns the abstract concept of emergence into a theoretically grounded reality for scaling populations of AI agents.

1 Introduction

Agentic systems based on foundation models are rapidly moving from single-agent chat assistants to multi-agent populations of autonomous agents that plan, communicate, use tools and act in the world. This shift is fueled by the expectation that scaling the number of agents will unlock performance gains analogous to the scaling laws of large language models Qian et al. [2025], Chaudhari et al. [2025]. However, analytical studies grounding such intuitions to the asymptotic behavior of these multi-agent systems remains limited. Current deployments, including AI-driven social networks like Moltbook Marzo and Garcia [2026] and autonomous personal assistants such as OpenClaw Wang et al. [2026] , treat agentic coordination as an engineering challenge to be addressed with more capable models, efficient communication, or improved prompts. We argue instead that coordination at scale is fundamentally an algorithmic challenge, one with a principled solution class: *mean-field learning*. Without it, systems may appear to work for small populations but become brittle as populations grow.

When foundation model (FM) agents interact, each agent's optimal behavior depends on the behaviors of its peers (the tools they use, the subtasks they claim, and the information they produce/consume) Tran et al. [2025]. In an N-agent system, tracking these dependencies exactly requires reasoning over a joint state-action space that grows combinatorially in N. The sample complexity of learning good collective behavior in this space, and the computational complexity of computing coordinated policies scale exponentially with N Blondel and Tsitsiklis [2000]. The multi-round debate, committee, and tournament patterns popular in today's agentic frameworks impose at least superlinear communication overhead and are structurally mismatched to the large-N

regime their proponents envision. Without a principled alternative, agentic multi-agent systems will hit a hard algorithmic wall before the promises of collective capabilities can be realized.

We argue that mean-field learning provides a principled alternative. The mean-field approach replaces explicit dependence on all other agents' states and actions with a compressed population-level signal (the empirical distribution of agent states) that serves as a sufficient statistic for each agent's decision problem. Applying mean-field theory to FM agents, such as LLM-based agents, is not a straightforward translation of existing multi-agent reinforcement learning (MARL) results Yang et al. [2018b], Guo et al. [2021]. Such agents differ fundamentally: their policies are expressed in natural language rather than tabular or neural action spaces, their coordination relies on semantic communication rather than numeric signals, and deployment typically involves pre-training followed by fine-tuning rather than reinforcement learning from scratch, which yields a very different sample-efficiency profile. We show that the mean-field framework extends naturally to accommodate these differences, giving a formal treatment of agentic emergence through the threshold phenomenon by which population-based collective capabilities become reliable as agent population size grows while individual agent capability is held fixed. This framework applies to homogeneous populations of identical agent (FM / LLM) instances, as well as heterogeneous populations combining orchestrator models, specialized fine-tuned executors, and domain-specific tools.

We ground our position in domains where the tension between expected scaling gains and coordination tractability is most acute. In **agentic coding**, large populations of LLM agents collaborating on shared codebases face tool contention, duplicated effort, and cascading dependencies that naive pairwise coordination cannot manage at scale; A mean-field signal may summarize task progress, tool use, and code quality to enable routing and load balancing without centralized management. In **robotics**, the mean-field framework formalizes and extends the population-level conditioning mechanisms already present in systems such as Gemini Robotics 1.5 Team et al. [2025] and $\pi_{0.7}$ Intelligence et al. [2026], providing the theoretical grounding those systems currently lack and showing how their population-level abstractions should be generalized to deployed multi-robot fleets. In each domain, the coordination bottleneck that limits scaling is precisely the problem mean-field theory was built to solve, and the population-level abstractions that the most capable systems already employ (compressed behavioral summaries, hierarchical planning-execution decompositions, etc.) are implicit mean-field reasoning waiting to be formalized.

2 Background

Agentic Scaling. With the development and large-scale deployment of LLM-based AI agents has come a sustained effort to scale agentic systems in the number of agents. Researchers and practitioners are overcoming the limited functionality of individual agents with interconnected networks of agents. Recent work focuses on improving the orchestration of multi-agent architectures through planning, state management, quality control, and standardized coordination protocols [Adimulam et al., 2026], as well as on improving dynamic topology routing to manage communication overhead while producing interpretable coordination traces [Lu et al., 2026]. The growing capabilities of these agents have motivated research in simulating large-scale agentic systems, such as environments capturing the social-media interactions of over 74,000 agents [Fadaei et al., 2026]. As agent populations grow, systems must accommodate non-uniform and possibly unreliable newcomers by proactively grouping tasks to familiarize them, leading to monotonic performance improvements [Shao et al., 2026, Chaudhari et al., 2025, Anand and Liaw, 2025]; similarly, they extend tool-use capabilities to control contexts available to agents [Gupta et al., 2026], and adapt decisions based on interaction experience across domains such as medicine and robotics [Pati, 2025].

These advances have inspired the vision of open-ended generalizable autonomous agents to assist scientists and engineers. Yet as agent populations grow, the need to understand their emergent behavior becomes acute, often left elusive due to a lack of interpretation of agentic interactions. This gap motivates our central position: the brute-force heuristics currently used to scale agent systems must give way to mathematically rigorous orchestration with multi-agent mean-field learning.

Mean-Field Multi-Agent Reinforcement Learning. MARL has long studied scaling challenges due to the exponential growth of the joint state-action space with the number of agents N [Blondel and Tsitsiklis, 2000]. Mean-field methods address this by replacing the intractable joint space with a population-level aggregate distribution over individual states and actions [Yang et al., 2018b, Carmona et al., 2023, Li et al., 2021, Guo et al., 2021, Chen et al., 2021, Cui et al., 2024, Gu et al., 2025, Lin et al., 2023]. Under homogeneity assumptions, joint policy optimization can be reduced to a problem of polynomial complexity whose approximation error vanishes at the rate $O(1/\sqrt{N})$ [Yang et al., 2018a, Gu et al., 2021]. This guarantee is the foundation of our position that mean-field methods convert unscalable problems into tractable ones.

A series of algorithmic advances has made this theoretically sound approach practical. Mean-field sampling techniques treat agents as part of a latent distribution and enable efficient cooperative RL without explicit pairwise interactions [Anand and Qu, 2024, Anand et al., 2025, 2026, Anand and Karmarkar, 2026]. Importantly, the homogeneity assumption can be relaxed. Graphon mean-field games (GMFGs) model heterogeneous populations by encoding non-uniform interactions via a graphon, which is a limiting object that captures a diversity of agent types and their couplings [Subramanian et al., 2022, Cui et al., 2023, Cui and Koeppl, 2022, Caines and Huang, 2018]. Recent work also extends GMFGs to sequential settings, sparse graphs, and learning from sampled data, while cooperative graphon mean-field control directly optimizes population-weighted joint objectives [Hu et al., 2023, Fabian et al., 2023, Zhang et al., 2023, Fabian et al., 2025].

These works demonstrate that mean-field theory scales to the kind of heterogenous, structured populations that arise in real-world LLM-agent deployments: the success of these methods does not depend on the specific policy representation, i.e. it can be tabular, neural, or LLM-based, but the core insight is the same: a compact population signal to replace exhaustive mutual modelling.

3 Mean-Field LLM Agents

We start by formalizing the concept of LLM agents and multi-agent systems, using the formulations in Guo et al. [2024] and Kim et al. [2025] as our foundation.

Definition 3.1 (Agents) *An LLM agent a_i is defined as a tuple $(\Phi_i, \mathcal{A}_i, M_i, \pi_i)$, where Φ_i is the LLM's reasoning policy, $\mathcal{A}_i = \{\mathrm{ToolCall}(t, \theta) : t \in \mathcal{T}, \theta \in \Theta_t\}$ is the action space of tool calls, where $\mathcal{T}$ is the set of available tools and Θ_t the valid parameters, M_i is the internal memory, and $\pi_i : \mathcal{H}_i \to \mathcal{A}_i$ is the decision function mapping observation histories to actions.*

Definition 3.2 (Agentic system) *An agentic system $\mathcal{S} = (A, E, C, \Omega)$ comprises a set of agents $A = \{a_1, \ldots, a_N\}$ for $N \geq 1$, a shared environment E, a communication topology C, and an orchestration policy Ω. Each agent a_i operates within the shared environment via iterative feedback.*

Consider a population of N agents in a shared space, as in Mi et al. [2025]. At time t, a subset of N_t agents ($\cup_t N_t = \{1, \ldots, N\}$) act simultaneously based on an encoding of personal attributes and observations of the environment, where N_t is the set of active agents at time t. Letting $c_t \in \{1, \ldots, N\}$, the collective $\{c_t\}_t$ can be used to define hierarchical or compositional agents [Du et al., 2023]. Next, the agent states are given by $\vec{s}_t = (s_t^1, \ldots, s_t^{N_t})$ with $s_t^i \in \mathcal{X}$, and $\vec{a}_t = (a_t^1, \ldots, a_t^{N_t})$ denotes the joint action vector, where each action influences the environment and future behavior.

3.1 Mean-Field AI Agent Framework

Simulating direct interactions between every pair of agents becomes infeasible as the number of agents N and the time horizon grow. While full interaction histories can be stored for small groups, they do not scale to large populations. Mean-field methods address this by modeling the collective as a whole [Mi et al., 2025]. Rather than tracking each agent individually, the mean-field maintains a single population-centric signal that evolves over time and summarizes the state of the group. This is initialized as an empty context m_0 and recursively updated by a dedicated mean-field model $m_t \leftarrow \mu(m_{t-1}, \vec{s}_{t-1}, \vec{a}_{t-1})$, where $(\vec{s}_{t-1}, \vec{a}_{t-1})$ are the ensemble of agent states and actions. The update distills the most decision-relevant information from the latest population trajectory.

Decision-Making. At each step, agent i receives its private state s_t^i and the current mean-field signal m_t, then selects an action by sampling from its LLM-based policy $a_t^i \sim \pi^i(\cdot|s_t^i, m_t)$. Since the policy operates over the unbounded space of natural language, it can produce highly adaptive and context-sensitive actions [Mi et al., 2025, Anand and Liaw, 2025]. Once all N_t agents act, the system transitions via $\vec{s}_{t+1} \sim P(\cdot|\vec{s}_t, \vec{a}_t, m_t)$, where P captures the dynamics of the shared world.

Mean-Field Learning. The mean-field signal should be a compact predictor of future agent behavior. We formalize this by applying the information-bottleneck principle [Mi et al., 2025]. Let $Y = \{a_t^{s_t}\}_{t=1}^{N_t}$ and $X = \{(\vec{s}_\tau^{*}, \vec{a}_\tau^{s_\tau})\}_{\tau=0}^{t}$ be the history. Then, the IB objective is $\min_{m_t} I(m_t; X) - \beta \cdot I(m_t; Y)$, where the first term encourages compression of the past and the second term preserves predictive information about upcoming actions. In practice $I(m_t; Y)$ is approximated by the log-likelihood of observed actions under the policy $\sum_{i=1}^{N_t} \log \pi(a_t^{*i} | s_t^i, m_t)$, and the compression term $I(m_t; X)$ is upper-bounded using a variational posterior $\mu_\phi(m_t|X)$ and a fixed prior $r(m_t)$, then $I(m_t; X) \leq \mathbb{E}_{p(X)}[\mathrm{KL}(\mu_\phi(m_t|X) \| r(m_t))]$. Setting $X = \{m_{t-1}, \vec{s}_{t-1}, \vec{a}_{t-1}\}$, for $m_t \sim \mu_\phi(\cdot | m_{t-1}, \vec{s}_{t-1}, \vec{a}_{t-1})$, balances predictive richness and efficiency, so the combined loss is

$$\mathcal{L}_{\text{mean-field}} = \sum_{t=1}^{T} \left(\mathbb{E}_{p(X_t)} D_{KL}(\mu_\phi(m_t|X_t) \| r(m_t)) - \beta \sum_{i=1}^{N_t} \log \pi(a_t^{*i} | s_t^i, m_t) \right). \tag{1}$$

Policy Optimization. The agent policy π is shared across the population and assumes a factorized form $\pi(\vec{a}_t | \vec{s}_t, m_t) = \prod_{i=1}^{N_t} \pi(a_t^i | s_t^i, m_t)$, where individual behavior differences arises naturally from the personalized language-expressed state inputs s_t^i, expressed in natural language. The policy is trained by minimizing the negative log-likelihood of ground-truth actions

$$\mathcal{L}_{policy} = -\sum_{t=1}^{T} \sum_{i=1}^{N_t} \log \pi(a_t^{s_t} | s_t^i, m_t), \tag{2}$$

which aligns the policy's output with empirically observed decisions conditioned on both private state and the population summary.

Heterogeneity. When agents differ in morphology or role, a single population-wide mean field may be insufficient. Here, the framework naturally generalizes to multi-class mean fields, where each agent type has its own aggregate signal, and cross-class coupling is captured by graphon mean-fields [Caines and Huang, 2021, Cui and Koeppl, 2022, Anand et al., 2026]. Motion transfer across embodiments in robotics (Appendix A) implicitly instantiate this multi-class structure.

4 Multi-Agent Emergent Mean-Field LLM Agents

Having defined the mean-field orchestration mechanism, we now turn to the phenomenon it is designed to explain and control: agentic emergence. The literature on emergence in large language models [Berti et al., 2025, Arora and Goyal, 2023] indicates that new capabilities often appear discontinuously as model scale increases. However, quantifying emergence is fraught with difficulty: evaluation metrics can create the illusion of sharp transitions [Schaeffer et al., 2023], different skills may emerge simultaneously, and the very notion of a "skill" in language agents is hard to formalize.

In the multi-agent LLM setting, we are interested in scaling the population of interacting agents while keeping per-agent capabilities fixed [Riedl, 2026]. A collective behavior, such as reliably executing a large-scale software migration or generalization across robot embodiments, may appear only when the number of agents N surpasses some threshold κ. Intuitively, any collective capability may have an intrinsic minimal population size. We might speak of a behavior being κ-emergent if it reliably manifests only when $N \geq \kappa$, where κ depends on the task complexity, the structure of the environment, and the coordination mechanism. For instance, a collective being moral might require at least two agents (via a debate), so morality might be a 2-emergent behavior. Similarly, expert coding agents can be κ-Emergent where κ is a function of difficulty of understanding a certain codebase. Using this notion, we define agentic emergence below:

Agentic Emergence. Consider an N-agent partially observable game of horizon T. Agent i has history $h_{i,t} \in \mathcal{H}_t$ and acts via a fixed policy $a_{i,t} \sim \pi_\theta(\cdot | h_{i,t}, u_t)$, where u_t is an optional broadcast signal. Crucially, the policy class π_θ does not grow in capability with N. We are interested in capabilities that arise only by increasing N, not from improving individual agents. Let $g : \mathcal{H}_t \to \mathcal{X}$ be a microstate extractor, with $x_{i,t} = g(h_{i,t})$. Define the empirical population measure

$$\mu_t^{(N)} = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_{i,t}}, \tag{3}$$

which summarizes the distribution of agent states at time t. Let $Z^{(N)} \in \{0, 1\}$ denote a macroscopic success event, e.g. discovering a valid hypothesis or completing a large software migration.

Definition 4.1 (Agentic emergence) *Fix a per-agent policy class π_θ, an N-agent environment family, and a macroscopic success event $Z^{(N)}$. We say that $Z^{(N)}$ is* agentically emergent *if there exist $\delta \in (0, 1/2)$, $c \in (0, 1)$, and a threshold $\kappa = N_\star(\delta)$ such that*

$$\sup_{N \leq cN_\star(\delta)} \Pr_{\pi_\theta}^{(N)}[Z^{(N)} = 1] \leq \delta, \qquad \inf_{N \geq N_\star(\delta)} \Pr_{\pi_\theta}^{(N)}[Z^{(N)} = 1] \geq 1 - \delta. \tag{4}$$

Thus, the same per-agent capability cannot reliably realize the macroscopic behavior below a population threshold, but can realize it reliably beyond that threshold. This isolates population-driven emergence from improvements in individual models by fixing π_θ and defining emergence as a reproducible, thresholded transition in collective performance.

Necessity of Mean-Field Learning. A naïve N-agent system conditions on the full joint state $\mathbf{x}_t = (x_{1,t}, \ldots, x_{N,t})$ and action $\mathbf{a}_t = (a_{1,t}, \ldots, a_{N,t})$. With d possible microstates and k possible actions per agent, the joint space scales as $d^N k^N$, which is a barrier to explicit reasoning. In contrast, policies of the form $a_{i,t} \sim \pi_\theta(\cdot|h_{i,t}, m_t)$ where $m_t = \phi(\mu_t^{(N)})$, replaces dependence on all other agents with dependence on the population distribution, changing the scaling from exponential to polynomial and making reasoning over the collective tractable.

Emergent Mean-Field Learning. The mean-field m_t should retain predictive information on future states or macroscopic success, while discarding inapt individual detail, yielding an objective:

$$\min_\phi \sum_{t=1}^{T} \left[\mathcal{L}_{\text{pred}}(m_t; Z^{(N)}, \mu_{t+1}^{(N)}) + \lambda \mathcal{L}_{\text{comp}}(m_t; \mu_t^{(N)}) \right]. \tag{5}$$

Given m_t, the agent policy can be trained via reinforcement learning [Rafailov et al., 2024], using:

$$\mathcal{L}_{\text{policy}}(\theta) = -\sum_{t=1}^{T} \sum_{i=1}^{N} \log \pi_\theta(a_{i,t}^\star \mid h_{i,t}, m_t), \quad J_N(\theta, \phi) = \mathbb{E}\left[\sum_{t=0}^{T} R(\mu_t^{(N)}, m_t, \mathbf{a}_t)\right]. \tag{6}$$

The mean-field dynamics/rewards decouple across agents

$$P(x_{i,t+1}|\mathbf{x}_t, \mathbf{a}_t) \approx P_{\text{MF}}(x_{i,t+1}|x_{i,t}, a_{i,t}, \mu_t^{(N)})$$

and $r_i(\mathbf{x}_t, \mathbf{a}_t) \approx r_{\text{MF}}(x_{i,t}, a_{i,t}, \mu_t^{(N)})$, moving the dependence on N from the joint state-action space into the estimation of a population statistic, thereby enabling tractable N-emergence behaviors.

5 Practical Implications of Mean-Field LLMs

5.1 Planning agents in Coding

Agentic coding tools, such as OpenAI Codex, Claude Code, Windsurf, and Cursor, are reshaping software engineering [Sapkota et al., 2025, AlOmar et al., 2024, Tufano et al., 2024]. Software engineering is a natural testbed for mean-field agentic coordination, since the environment is structured, heavily instrumented, and deeply coupled [Bellur et al., 2025, Shirafuji et al., 2023, DePalma et al., 2024]. A modern codebase is not merely a collection of files; it is a shared dynamical system consisting of modules, dependency graphs, build configurations, and deployment constraints. Agents operating in this environment must coordinate over both code semantics and shared resources, such as which files are being edited, which tests are failing, which dependencies are unstable, and which CI/review bottlenecks are binding. While adding more agents increases parallelism, it also amplifies duplicated work, conflicting patches, and CI congestion [Li et al., 2025]; however, explicit pairwise communication, where each agent reads all others' plans and negotiates task allocation, can resolve conflicts in small teams, but scales quadratically in communication overhead.

This perspective illuminates some challenges: in agentic refactoring, failures often stem from incomplete global context and missed dependencies [Horikawa et al., 2025]. In the mean-field view, these are population-state estimation failures: the system lacks a reliable aggregate signal describing which invariants have been checked, which modules remain inconsistent, and which dependencies have already been migrated. Similarly, decentralized LLM collaboration methods [Liu et al., 2026] can be interpreted as learning policies for agents that coordinate through shared signals rather than exhaustive pairwise deliberation, which is an implicit step toward mean-field LLMs.

5.2 Physical embodied agents in Robotics

Robotic multi-agent systems present a demanding instantiation of FM / LLM agents. In addition to coarse natural language for representing state and action spaces, embodied agents must also ground population-level coordination in physical dynamics: shared workspace constraints, real-time kinematic coupling, and heterogeneous morphologies that induce structurally distinct action spaces within the same fleet [Fung et al., 2025, Lin et al., 2024, Preiss et al., 2017]. These challenges expose both the necessity and the limits of naive multi-agent scaling, and we argue that mean-field learning provides the only principled pathway to coordination at realistic fleet sizes.

We support this argument in Appendix A by examining two independently developed foundation model frameworks for physical robotics: (i) the Gemini Robotics 1.5 framework [Team et al., 2025], comprising the VLA model GR 1.5 and the VLM orchestrator GR-ER 1.5, and (ii) the $\pi_{0.7}$ framework from Physical Intelligence, comprising the Hi Robot hierarchical inference framework [Shi et al., 2025] and a generalist VLA model [Intelligence et al., 2026].

Our position is that both the cross-embodiment generalization using Motion Transfer in Gemini Robotics 1.5, and the compositional task generalization shown by $\pi_{0.7}$, are empirical validations of an implicit mean-field formalism, albeit along two complementary axis. In the case of Motion Transfer, the learned mean field representation balances the compression versus prediction information bottleneck in the space of different robot morphologies. On the other hand, to achieve compositional generalization in the space of task description in $\pi_{0.7}$, the mean field representation optimizes the physical grounding of natural language prompts. Neither outline any principled exploration of scaling their agentic frameworks, therefore our analysis in Appendix A attempts to ground their empirical observations in the emergent mean-field learning framework and provides insights into potential avenues for future research.

6 Illusion of Scaling Laws for Emergence

Much of the practical interest in AI Agents (as in Section 5) stems from ideas of *emergence*: that scaling interactions between agents can often unlock new problem-solving capabilities. A critical study of recent literature reveals that this narrative is largely an illusion due to computational tractability challenges: when analyzed by applying game-theoretic lower bounds and information-theoretic limits, and looking at large-scale controlled ablations, the assumption that scaling the number of agents is “all that you need" dissipates [Li et al., 2024].

6.1 The Measurement Artifact: The Mirage of Discontinuous Capability Jumps

The assumption driving the search for “emergent abilities" through population scaling is derived from emergence in base AI models, where capabilities suddenly arise once a certain parameter count or compute threshold is reached. It is also intuitive in real world problems where specialized groups of N agents coordinate with each other to solve complicated tasks. If individual models demonstrate this new behavior, prevailing heuristics suggest that similar discontinuous jumps in ability should occur in collectives as N increases.

Consider the following complex task: A model (we assume it to be a specialized agent) is asked to output a sequence of L tokens exactly without error. Doing this successfully results in a reward of 1; otherwise the reward is 0. Here, the probability of getting an exact match scales exponentially as p^L, where p is the model's per-token accuracy. A smooth improvement with p as the model scales will inevitably produce an exponentially-delayed spike in task performance, giving the illusion of emergent capabilities. When continuous linear metrics such as Token Edit Distance or Brier Score [Brier, 1950] are used to evaluate these identical outputs, the discontinuous jump in ability vanishes to reveal a smooth scaling law. Our position is that N agents when giving an output of L tokens each will retain a similar scaling law that grows exponentially with more N and L.

This “mirage” is significant for multi-agent systems: the perceived success of a large debate or voting ensemble is almost exclusively the result of a discontinuous evaluation metric amplifying a marginal fractional increase in the underlying probability of success, rather than the collective agents suddenly unlocking a genuinely novel cognitive algorithm. Relying on this pervasive illusion encourages practitioners to endlessly scale population counts, operating under the false assumption that a massive ensemble will eventually cross a hidden threshold for complex tasks.

6.2 Game-Theoretic Limits: The Sample Complexity of Artificial Social Rationality

Beyond mere measurement artifacts, the failure of naive scaling of agents is firmly rooted in fundamental, mathematically proven sample complexity limits. A widespread assumption in system design is that if individual agents are highly capable (i.e., individually rational), their interactions from their unstructured collective will converge on optimal cooperation. Bandyopadhyay et al. [2025] weakens this assumption within the framework of socially rational populations.

Consider a two-player finitely repeated general-sum game heavily obscured by private types of agents, defined by the tuple $\mathcal{G} = (\mathcal{I}, \mathcal{A}, \Theta, G, T)$. One of these agents can be selected from a population of agents at scale, inspired by replicator dynamics Börgers and Sarin [1997], quantifying the proportion of each type of agents in a population. Agents operate with strictly private utility functions dictated by their hidden type $\theta \in \Theta$. Utilities signify the reward or incentive for agents to do a task while type is a model of the agents' behavior making all actions and resultant payoffs heavily context-dependent. The target population is defined mathematically as "socially intelligent" if and only if it satisfies two conditions: (i) **Consistency**: Individual agents are Hannan-consistent (the no-regret property), meaning their external regret $\frac{1}{T}R_i^{ext}(h_T;\theta) \leq \epsilon$ is tightly bounded with high probability, and (ii) **Compatibility**: The paired agents successfully achieve an expected utility comparable to a Pareto-Optimal Nash Equilibrium (PONE) of the underlying game.

The critical question is whether an uninitiated algorithm can achieve zero-shot cooperation when abruptly introduced into this highly structured, socially intelligent population. The learning objective here is to minimize *Altruistic Regret* R_T, which measures the cumulative deficit between the partner's realized payoffs and their lowest-payoff PONE at time T [Bandyopadhyay et al., 2025].

In general, individual consistency and compatibility are fundamentally insufficient for zero-shot cooperation of agentic populations at scale. To illustrate the sample complexity inherent in agentic cooperation, we examine lower bounds in Bandyopadhyay et al. [2025] which show that for any $\delta, \epsilon > 0$, there exists a class of games and a class of (δ, ϵ, T)-socially intelligent agents such that, for any data-dependent meta-strategy informed by a dataset of K interaction histories, the expected altruistic regret is bounded from below: for any data-dependent meta-strategy informed by K interaction histories, the expected regret $\mathbb{E}[R_T]$ scales as:

$$\mathbb{E}[R_T] \geq \Omega\left((T-L)\min\left\{\frac{1}{2}, \frac{N^{L-2}-1}{2K+1}\right\}\right) \tag{7}$$

for some required authentication sequence length L.

Here, authentication means that the agents have to ensure that their collaborative partner does not deviate from the optimal strategy, thereby protecting from adversarial agents which try to pose as a member of the collective. This lower bound shows that the sample complexity required for cooperative behavior grows exponentially with the interaction history (N^{L-2}). This formally proves the established intractability of the $d^N k^N$ joint state-action space from Section 4.

To avoid this combinatorial explosion in complexity, Bandyopadhyay et al. [2025] gives an upper bound showing that the exponential lower bound can be circumvented if the agent successfully models the aggregate behavior of the interactive population over a short horizon $\tilde{T}$. Though their proof is applied to create a static "imitate-then-commit" meta-strategy, we argue this same population-level dynamic can be addressed adaptively using mean-field MARL. Instead of imitation learning, we capture this aggregate behavior continuously through the mean-field signal m_t. Using the Information Bottleneck defined in Section 3.1, we actively compress the interaction history. The expected altruistic regret under this approach satisfies:

$$\mathbb{E}[R_T] \leq \delta(K) + \epsilon_1 + \delta_1 + \frac{T-\tilde{T}}{T}(\epsilon_0 + \delta_0) \tag{8}$$

where $\delta(K)$ decays smoothly and polynomially with the dataset size K, and ϵ_i, δ_i are error bounds and failure probabilities for $i \in \{0, 1\}$ signifying consistency and compatibility, respectively Bandyopadhyay et al. [2025].

The implications for scaling in Equations 7 and 8 are absolute: the lower bound proves that scaling a large number of "rational" agents into a shared environment will never spontaneously yield emergent cooperation due to the exponential compute barrier of hidden types and complex histories. In fact, the upper bound proves that sustained success strictly requires analyzing the system at the population level by learning the conventions and statistical behaviors of the aggregate group rather than exhaustively attempting to model individual peers. This validates our position from Section 4: true agentic emergence demands that policies be conditioned on the empirical population measure $\mu_t^{(N)}$. This way, we shift the burden of coordinating from the exponential complexity of joint state-action space modeling to the far more tractable polynomial estimation at the population level.

7 Alternative Views

Our position is that scaling agentic systems require population-level abstractions, and mean-field learning is the most principled current framework for designing such abstractions. However, there are several credible alternative views. We address them here to clarify the scope of our claim.

A.1: Direct Systems Optimization May Be Enough. One view is that agent-scaling bottlenecks are engineering problems, solvable by faster inference and cheaper accelerators, rendering new frameworks unnecessary. We agree that system optimization is vital: many near-term gains in agent deployment will come from these improvements. However, it does not remove the underlying interaction problem. Improving per-agent compute does not change the asymptotic structure of explicit multi-agent planning. A system that requires each agent to track every other agent's state remains structurally mismatched to the large-N regime even if each inference call is free. The distinction we draw is between making the current paradigm cheaper and changing its scaling law. Mean-field methods address the latter by replacing explicit dependence on individual agents with one on an aggregate population state. It is not a substitute for systems optimization, but a complementary abstraction that determines what the system should compute as N grows.

A.2: Token Efficiency May Matter More Than Algorithmic Efficiency. Another view is that the dominant cost in LLM-agent systems is token use, so reducing prompt length and training models to use fewer deliberation tokens may make multi-agent systems practical without special orchestration. While this certainly lowers reasoning costs, multi-agent scaling introduces costs that are irreducible to agent token counts: communication fan-out, conflicting actions, and resource congestion. A perfectly token-efficient agent can still be in an inefficient collective if the collective lacks stable population-level control. Scalable systems require both token compression and population-state compression. Mean-field learning provides the latter: instead of each agent reading long peer transcripts, the system compresses collective behavior into a sufficient statistic.

A.3: Hierarchical Orchestration Could Replace Mean-Field Learning. A third view is that stronger orchestrators (hierarchies of planners, reviewers, and executors) can decompose tasks and resolve conflicts without population-level abstractions, a pattern already common in coding agents. Hierarchy is often the right architecture; however, the critical question is what information flows through it. A top-level orchestrator that explicitly tracks the state of lower-level agents still faces the same scaling problem: as N grows, the orchestrator becomes a centralized bottleneck whose context must either expand with N or discard information in an unprincipled way. Mean-field learning can resolve this by abstracting the missing information, where the upper layer receives population summaries, and the lower layer receives aggregate signals.

A.4: The Measurement Problem. One might argue that the field moves too fast for asymptotics to be useful: models, tools, and deployment practices change rapidly, and gains from multi-agent systems are often confounded by stronger base models, better prompts, larger context windows, or larger evaluation budgets. In this lens, measurement is the urgent bottleneck rather than the theory. Indeed, many reported improvements in agent systems do not isolate the contribution of multi-agent interactions from simply using more tokens, more wall-clock time, or a better judge. This is a valid concern, and we structure our call to action for the larger community to resolve this.

8 Call to Action

Current practice often evaluates agent systems by asking whether adding more agents improves a task metric. We believe this is insufficient for understanding when additional agents help, when they interfere, and which abstractions allow the system to remain controllable as N grows. Thus, we call for an agenda centered on mean-field agentic scaling.

A1: Researchers. Researchers should report agentic emergence as a scaling phenomenon under a fixed agent capability, specifying the base policy, population size, and interaction topology. Without these, it is difficult to distinguish genuine emergence from other factors. In addition to outputs and transcripts, researchers should log population states by recording task stage, tool calls, communication events, and resource use to estimate microstate distributions and broadcast signals, enabling direct tests of whether a mean field signal improves collective outcomes. Researchers should also develop scaling-aware theoretical guarantees to explicitly track how the mean-field approximation error and communication cost depend on population size, heterogeneity, and topology.

A2: Practitioners. Practitioners should instrument population states early by exposing aggregate metrics (which tasks agents are working on, where failures concentrate, how often work is repeated, and how traffic grows with N). Secondly, they should avoid scaling by all-to-all deliberation: multi-round debate and committee voting may suffice at small scale, but before adding agents, practitioners should examine how communication, evaluation, and tool-use costs scale with N. Finally, deployment teams should monitor for negative emergence (synchronous mistakes, correlated hallucinations, duplicated tool calls, and collective overconfidence) which are pathologies invisible in small-scale experiments, requiring population-level diagnostics and interventions.

A3: For the Broader Community. The broader AI agents research community should build more datasets to study agent behavior at the population level. Many contemporary agent traces are kept private, are too task-specific, or cannot be collectively analyzed. Public datasets need multi-agent trajectories with tool calls and intermediate states that cover a wide variety of domains, including coding and embodied control. Moreover, the community should also set a higher standard on the reporting norms for agentic scaling. Papers should report not only the number of agents used, but also the communication topology, total token and per-agent budgets, and marginal benefit of each additional agent. A claim that a system scales should include evidence about how cost and performance change with N, not merely that a larger system outperforms a smaller one.

References

Apoorva Adimulam, Rajesh Gupta, and Sumit Kumar. The Orchestration of Multi-Agent Systems: Architectures, Protocols, and Enterprise Adoption, 2026. URL https://arxiv.org/abs/2601.13671.

Eman Abdullah AlOmar, Anushkrishna Venkatakrishnan, Mohamed Wiem Mkaouer, Christian Newman, and Ali Ouni. How to refactor this code? An exploratory study on developer-ChatGPT refactoring conversations. In *Proceedings of the 21st International Conference on Mining Software Repositories*, MSR '24, page 202–206, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400705878. doi: 10.1145/3643991.3645081. URL https://doi.org/10.1145/3643991.3645081.

Emile Anand and Ishani Karmarkar. Learning Approximate Nash Equilibria in Cooperative Multi-Agent Reinforcement Learning via Mean-Field Subsampling, 2026. URL https://arxiv.org/abs/2603.03759.

Emile Anand and Sarah Liaw. Feel-Good Thompson Sampling for Contextual Bandits: a Markov Chain Monte Carlo Showdown, 2025. URL https://arxiv.org/abs/2507.15290.

Emile Anand and Guannan Qu. Efficient Reinforcement Learning for Global Decision Making in the Presence of Local Agents at Scale. *arXiV*, 2024.

Emile Anand, Richard Hoffmann, Sarah Liaw, and Adam Wierman. Graphon Mean-Field Subsampling for Cooperative Heterogeneous Multi-Agent Reinforcement Learning, 2026. URL https://arxiv.org/abs/2602.16196.

Emile Timothy Anand, Ishani Karmarkar, and Guannan Qu. Mean-Field Sampling for Cooperative Multi-Agent Reinforcement Learning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=CTsdZ3j6dR.

Sanjeev Arora and Anirudh Goyal. A theory for emergence of complex skills in language models. *arXiv preprint arXiv:2307.15936*, 2023.

Saptarashmi Bandyopadhyay, Mustafa Mert Çelikok, and Robert Loftin. On the Complexity of Learning to Cooperate in Populations of Socially Rational Agents. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, pages 233–241, 2025.

Abhiram Bellur, Fraol Batole, Mohammed Raihan Ullah, Malinda Dilhara, Yaroslav Zharov, Timofey Bryksin, Kai Ishikawa, Haifeng Chen, Masaharu Morimoto, Shota Motoura, Takeo Hosomi, Tien N. Nguyen, Hridesh Rajan, Nikolaos Tsantalis, and Danny Dig. Leveraging LLMs, IDEs, and Semantic Embeddings for Automated Move Method Refactoring, 2025. URL `https://arxiv.org/abs/2503.20934`.

Leonardo Berti, Flavio Giorgi, and Gjergji Kasneci. Emergent abilities in large language models: A survey. *arXiv preprint arXiv:2503.05788*, 2025.

Vincent D. Blondel and John N. Tsitsiklis. A Survey of Computational Complexity Results in Systems and Control. *Automatica*, 36(9):1249–1274, 2000. ISSN 0005-1098. doi: https://doi.org/10.1016/S0005-1098(00)00050-9.

Tilman Börgers and Rajiv Sarin. Learning Through Reinforcement and Replicator Dynamics. *Journal of Economic Theory*, 77(1):1–14, 1997.

Glenn W Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950. doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

Peter E. Caines and Minyi Huang. Graphon Mean Field Games and the GMFG Equations. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 4129–4134, Miami Beach, USA, 2018. doi: 10.1109/CDC.2018.8619367.

Peter E. Caines and Minyi Huang. Graphon Mean Field Games and Their Equations. *SIAM Journal on Control and Optimization*, 59(6):4373–4399, January 2021. ISSN 1095-7138. doi: 10.1137/20m136373x.

René Carmona, Mathieu Laurière, and Zongjun Tan. Model-free mean-field reinforcement learning: Mean-field MDP and mean-field Q-learning. *The Annals of Applied Probability*, 33(6B):5334 – 5381, 2023. doi: 10.1214/23-AAP1949.

Shreyas Chaudhari, Srinivasa Pranav, Emile Anand, and José M. F. Moura. Peer-to-Peer Learning Dynamics of Wide Neural Networks. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, page 1–5, Hyderabad, India, April 2025. IEEE. doi: 10.1109/icassp49660.2025.10890126.

Minshuo Chen, Yan Li, Ethan Wang, Zhuoran Yang, Zhaoran Wang, and Tuo Zhao. Pessimism Meets Invariance: Provably Efficient Offline Mean-Field Multi-Agent RL. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, Virtual, 2021.

Kai Cui and Heinz Koeppl. Learning Graphon Mean Field Games and Approximate Nash Equilibria. In *International Conference on Learning Representations*, 2022. URL `https://openreview.net/forum?id=0sgntlpKDOz`.

Kai Cui, Christian Fabian, and Heinz Koeppl. Multi-Agent Reinforcement Learning via Mean Field Control: Common Noise, Major Agents and Approximation Properties, 2023. URL `https://arxiv.org/abs/2303.10665`.

Kai Cui, Christian Fabian, Anam Tahir, and Heinz Koeppl. Major-Minor Mean Field Multi-Agent Reinforcement Learning, 2024.

Kayla DePalma, Izabel Miminoshvili, Chiara Henselder, Kate Moss, and Eman Abdullah AlOmar. Exploring ChatGPT's code refactoring capabilities: An empirical study. *Expert Syst. Appl.*, 249 (PB), September 2024. ISSN 0957-4174. doi: 10.1016/j.eswa.2024.123602. URL `https://doi.org/10.1016/j.eswa.2024.123602`.

Yilun Du, Shuang Li, Antonio Torralba, Joshua B Tenenbaum, and Igor Mordatch. Improving factuality and reasoning in language models through multiagent debate. In *Forty-first International Conference on Machine Learning*, 2023.

Christian Fabian, Kai Cui, and Heinz Koeppl. Learning Sparse Graphon Mean Field Games, 2023.

Christian Fabian, Kai Cui, and Heinz Koeppl. Learning Mean Field Control on Sparse Graphs. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 15637–15660, Vancouver, Canada, 13–19 Jul 2025. PMLR.

Faezeh Fadaei, Jenny Carla Moran, and Taha Yasseri. Gender Dynamics and Homophily in a Social Network of LLM Agents. *arXiv preprint arXiv:2602.02606*, 2026.

Pascale Fung, Yoram Bachrach, Asli Celikyilmaz, Kamalika Chaudhuri, Delong Chen, Willy Chung, Emmanuel Dupoux, Hongyu Gong, Hervé Jégou, Alessandro Lazaric, Arjun Majumdar, Andrea Madotto, Franziska Meier, Florian Metze, Louis-Philippe Morency, Théo Moutakanni, Juan Pino, Basile Terver, Joseph Tighe, Paden Tomasello, and Jitendra Malik. Embodied AI Agents: Modeling the World, 2025. URL `https://arxiv.org/abs/2506.22355`.

Haotian Gu, Xin Guo, Xiaoli Wei, and Renyuan Xu. Mean-Field Controls with Q-Learning for Cooperative MARL: Convergence and Complexity Analysis. *SIAM Journal on Mathematics of Data Science*, 3(4):1168–1196, 2021. doi: 10.1137/20M1360700. URL `https://doi.org/10.1137/20M1360700`.

Haotian Gu, Xin Guo, Xiaoli Wei, and Renyuan Xu. Mean-Field Multiagent Reinforcement Learning: A Decentralized Network Approach. *Math. Oper. Res.*, 50(1):506–536, February 2025. ISSN 0364-765X. doi: 10.1287/moor.2022.0055.

Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*, 2024.

Xin Guo, Anran Hu, Renyuan Xu, and Junzi Zhang. Learning Mean-Field Games, 2021.

Karan Gupta, Pranav Vajreshwari, Yash Pandya, Raghav Magazine, Akshay Nambi, and Ahmed Awadallah. Scaling Agentic Capabilities, Not Context: Efficient Reinforcement Finetuning for Large Toolspaces, 2026. URL `https://arxiv.org/abs/2603.06713`.

Kosei Horikawa, Hao Li, Yutaro Kashiwa, Bram Adams, Hajimu Iida, and Ahmed E. Hassan. Agentic Refactoring: An Empirical Study of AI Coding Agents, 2025. URL `https://arxiv.org/abs/2511.04824`.

Yuanquan Hu, Xiaoli Wei, Junji Yan, and Hengxi Zhang. Graphon mean-field control for cooperative multi-agent reinforcement learning. *Journal of the Franklin Institute*, 360(18):14783–14805, 2023. ISSN 0016-0032. doi: https://doi.org/10.1016/j.jfranklin.2023.09.002.

Physical Intelligence, Bo Ai, Ali Amin, Raichelle Aniceto, Ashwin Balakrishna, Greg Balke, Kevin Black, George Bokinsky, Shihao Cao, Thomas Charbonnier, et al. $\pi_{0.7}$: A Steerable Generalist Robotic Foundation Model with Emergent Capabilities. *arXiv preprint arXiv:2604.15483*, 2026.

Yubin Kim, Ken Gu, Chanwoo Park, Chunjong Park, Samuel Schmidgall, A Ali Heydari, Yao Yan, Zhihan Zhang, Yuchen Zhuang, Mark Malhotra, et al. Towards a science of scaling agent systems. *arXiv preprint arXiv:2512.08296*, 2025.

Hao Li, Haoxiang Zhang, and Ahmed E. Hassan. The Rise of AI Teammates in Software Engineering (SE) 3.0: How Autonomous Coding Agents Are Reshaping Software Engineering, 2025. URL `https://arxiv.org/abs/2507.15003`.

Junyou Li, Qin Zhang, Yangbin Yu, Qiang Fu, and Deheng Ye. More Agents Is All You Need, 2024. URL `https://arxiv.org/abs/2402.05120`.

Yan Li, Lingxiao Wang, Jiachen Yang, Ethan Wang, Zhaoran Wang, Tuo Zhao, and Hongyuan Zha. Permutation Invariant Policy Optimization for Mean-Field Multi-Agent Reinforcement Learning: A Principled Approach, 2021.

Yiheng Lin, James A Preiss, Emile Timothy Anand, Yingying Li, Yisong Yue, and Adam Wierman. Online Adaptive Policy Selection in Time-Varying Systems: No-Regret via Contractive Perturbations. In *Thirty-seventh Conference on Neural Information Processing Systems*, New Orleans, USA, 2023.

Yiheng Lin, James A. Preiss, Fengze Xie, Emile Anand, Soon-Jo Chung, Yisong Yue, and Adam Wierman. Online Policy Optimization in Unknown Nonlinear Systems. In Shipra Agrawal and Aaron Roth, editors, *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 3475–3522, Edmonton, Canada, 30 Jun–03 Jul 2024. PMLR.

Shuo Liu, Tianle Chen, Ryan Amiri, and Christopher Amato. Learning Decentralized LLM Collaboration with Multi-Agent Actor Critic. *arXiv preprint arXiv:2601.21972*, 2026.

Yuxing Lu, Yucheng Hu, Xukai Zhao, and Jiuxin Cao. DyTopo: Dynamic Topology Routing for Multi-Agent Reasoning via Semantic Matching. *arXiv preprint arXiv:2602.06039*, 2026.

Giordano De Marzo and David Garcia. Collective behavior of ai agents: the case of moltbook, 2026. URL `https://arxiv.org/abs/2602.09270`.

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.

Ashis Kumar Pati. Agentic AI: a comprehensive survey of technologies, applications, and societal implications. *IEEE Access*, 2025.

James A. Preiss, Wolfgang Honig, Gaurav S. Sukhatme, and Nora Ayanian. Crazyswarm: A large nano-quadcopter swarm. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3299–3304, 2017. doi: 10.1109/ICRA.2017.7989376.

Chen Qian, Zihao Xie, YiFei Wang, Wei Liu, Kunlun Zhu, Hanchen Xia, Yufan Dang, Zhuoyun Du, Weize Chen, Cheng Yang, Zhiyuan Liu, and Maosong Sun. Scaling Large Language Model-based Multi-Agent Collaboration. In *The Thirteenth International Conference on Learning Representations*, 2025. URL `https://openreview.net/forum?id=K3n5jPkrU6`.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct Preference Optimization: Your Language Model is Secretly a Reward Model, 2024. URL `https://arxiv.org/abs/2305.18290`.

Christoph Riedl. Emergent Coordination in Multi-Agent Language Models, 2026. URL `https://arxiv.org/abs/2510.05174`.

Ranjan Sapkota, Konstantinos I. Roumeliotis, and Manoj Karkee. Vibe Coding vs. Agentic Coding: Fundamentals and Practical Implications of Agentic AI, 2025. URL `https://arxiv.org/abs/2505.19443`.

Rylan Schaeffer, Brando Miranda, and Sanmi Koyejo. Are Emergent Abilities of Large Language Models a Mirage?, 2023. URL `https://arxiv.org/abs/2304.15004`.

Shuai Shao, Yixiang Liu, Bingwei Lu, and Weinan Zhang. MonoScale: Scaling Multi-Agent System with Monotonic Improvement. *arXiv preprint arXiv:2601.23219*, 2026.

Lucy Xiaoyang Shi, Michael Robert Equi, Liyiming Ke, Karl Pertsch, Quan Vuong, James Tanner, Anna Walling, Haohuan Wang, Niccolo Fusai, Adrian Li-Bell, et al. Hi Robot: Open-Ended Instruction Following with Hierarchical Vision-Language-Action Models. In *Forty-second International Conference on Machine Learning*, 2025.

Atsushi Shirafuji, Yusuke Oda, Jun Suzuki, Makoto Morishita, and Yutaka Watanobe. Refactoring Programs Using Large Language Models with Few-Shot Examples. In *2023 30th Asia-Pacific Software Engineering Conference (APSEC)*, pages 151–160, 2023. doi: 10.1109/APSEC60848.2023.00025.

Sriram Ganapathi Subramanian, Matthew E. Taylor, Mark Crowley, and Pascal Poupart. Decentralized Mean Field Games, 2022. URL https://arxiv.org/abs/2112.09099.

Gemini Robotics Team, Abbas Abdolmaleki, Saminda Abeyruwan, Joshua Ainslie, Jean-Baptiste Alayrac, Montserrat Gonzalez Arenas, Ashwin Balakrishna, Nathan Batchelor, Alex Bewley, Jeff Bingham, et al. Gemini robotics 1.5: Pushing the frontier of generalist robots with advanced embodied reasoning, thinking, and motion transfer. *arXiv preprint arXiv:2510.03342*, 2025.

Khanh-Tung Tran, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O'Sullivan, and Hoang D. Nguyen. Multi-Agent Collaboration Mechanisms: A Survey of LLMs, 2025. URL https://arxiv.org/abs/2501.06322.

Rosalia Tufano, Antonio Mastropaolo, Federica Pepe, Ozren Dabic, Massimiliano Di Penta, and Gabriele Bavota. Unveiling ChatGPT's Usage in Open Source Projects: A Mining-based Study. In *Proceedings of the 21st International Conference on Mining Software Repositories*, MSR '24, page 571–583, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400705878. doi: 10.1145/3643991.3644918. URL https://doi.org/10.1145/3643991.3644918.

Yinjie Wang, Xuyang Chen, Xiaolong Jin, Mengdi Wang, and Ling Yang. Openclaw-rl: Train any agent simply by talking. *arXiv preprint arXiv:2603.10165*, 2026. URL https://arxiv.org/abs/2603.10165.

Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, and Jun Wang. Mean Field Multi-Agent Reinforcement Learning. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5571–5580. PMLR, 10–15 Jul 2018a. URL https://proceedings.mlr.press/v80/yang18d.html.

Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, and Jun Wang. Mean Field Multi-Agent Reinforcement Learning. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5571–5580, Stockholm, Sweden, 10–15 Jul 2018b. PMLR.

Fengzhuo Zhang, Vincent Y. F. Tan, Zhaoran Wang, and Zhuoran Yang. Learning regularized monotone graphon mean-field games, 2023.

A Mean Field Learning in Systems for Embodied AI Agents

In robotics, multi-agent systems can be used to model a team of individual robots, or even the components of a single robot (for example, a pair of robot arms) that jointly coordinate to execute a task. Practical applications of robotics in multi-agent settings, such as heterogeneous robot fleets or dexterous manipulators, expose both the necessity and the limits of naive scaling. Our position in favor of using mean field learning for principled scaling naturally finds support in the robotics setup. We lay out our arguments in this section by examining two independently developed foundation model families for physical robotics: Gemini Robotics 1.5 [Team et al., 2025] and $\pi_{0.7}$ [Intelligence et al., 2026].

Remark. We note that the following is our interpretation of the results presented in Team et al. [2025] and Intelligence et al. [2026] through the lens of mean-field learning. We call for further research to the conceptual directions outlined below, to further validate the empirical results reported by the authors of Gemini Robotics 1.5 and $\pi_{0.7}$.

A.1 The Embodied Agent System

We instantiate the agentic system $\mathcal{S} = (\mathbf{A}, E, \mathcal{C}, \Omega)$ of Definition 3.2 for a robotic fleet as follows. The environment E is a physical workspace with dynamics governed by rigid-body physics and shared resource constraints, such as workspace volume, sensor coverage, communication bandwidth. The agent set $\mathbf{A} = \{a_1, \ldots, a_N\}$ is a fleet of N robots. Following Definition 3.2, each agent $a_i = (\Phi_i, \mathcal{A}_i, M_i, \pi_i)$ instantiates a foundation model Φ_i, which can be a VLM, VLA, or world model, an action space $\mathcal{A}_i$ consisting of tool calls over motor primitives and perception APIs, a memory M_i encoding task history and sensorimotor traces, and a decision function $\pi_i : \mathcal{H} \to \mathcal{A}_i$ mapping observation histories to actions.

In this setting, the coordination bottleneck is immediate. In a fleet of N manipulation arms sharing a workspace, or N mobile robots engaged in collaborative coverage, each agent's optimal action depends on the current configurations of all other agents. Tracking the joint state $\vec{s}_t = (s_t^1, \ldots, s_t^N)$ exactly scales as $|\mathcal{S}|^N$, placing naive coordination firmly in the intractable regime identified in Section 4. The mean-field signal m_t, updated as $m_t \leftarrow \mu(m_{t-1}, \vec{s}_{t-1}, \vec{a}_{t-1})$ replaces this exponential joint state with a population-level aggregate, enabling each agent's policy $\pi_i(\cdot \mid s_t^i, m_t)$ to condition on collective behavior without explicit pairwise communication. In the robotic setting, m_t can be thought of as a compressed representation of system dynamics, agent kinematics, failure modes, or policy uncertainty.

The most structurally significant observation across both system families [Team et al., 2025, Intelligence et al., 2026] is their convergence on an identical two-level hierarchical architecture.

In Gemini Robotics 1.5, GR-ER 1.5 is a VLM which serves as a high-level orchestrator that performs spatial reasoning, multi-step task planning, progress estimation, and external tool use including web search. GR 1.5 is a VLA which serves as a low-level executor that translates perceptual observations and natural language instructions from the VLM orchestrator into motor commands, with its own internal Embodied Thinking capability that interleaves reasoning traces with action generation. GR-ER 1.5 can call GR 1.5 as a subagent, alongside any user-defined tool.

On the other hand, $\pi_{0.7}$ instantiates precisely the same decomposition under a dual-process framing: a System 2 high-level VLM that reasons through complex prompts and generates intermediate language subtasks, and a System 1 low-level VLA that executes those subtasks reactively. The convergence of both Team et al. [2025] and Intelligence et al. [2026] at this two-level hierarchical structure is strong evidence that it reflects a genuine inductive bias of physical task execution rather than a contingent design choice.

In a fleet of N robots, each robot runs one orchestration agent and one execution agent. The orchestration mean field m_t^{orch} summarizes the population distribution of planning states across the orchestration class. The execution mean field m_t^{exec} summarizes the population distribution of motor contexts and reasoning traces across the execution class. Each class conditions on the other's mean field rather than on the full joint state of all peer agents, recovering the polynomial sample complexity scaling established in Section 4.

This hierarchical nature also hints at the potential for agentic emergence in this setting following a Stackelberg mean-field formulation, in which the VLM class constitutes a population of leaders and the VLA class constitutes a population of followers, with the equilibrium concept incorporating the hierarchical dependency. Such a formulation would also capture the temporal asymmetry of embodied tasks, where high-level planning necessarily precedes and constrains low-level motor commitment, and where the cost of replanning is much higher than the cost of reactive execution adjustment. We believe this could be an important direction of future research in this space.

A.2 Population-Level Conditioning as Empirical Mean Field

Beyond the shared hierarchical architecture, both system families independently instantiate the mean-field signal m_t of Section 3 as a concrete engineering mechanism, though in different forms that together illuminate complementary aspects of the framework.

In Gemini Robotics 1.5, the Motion Transfer (MT) mechanism enables a single GR 1.5 checkpoint to learn from heterogeneous data collected across physically distinct robot platforms: ALOHA, Bi-arm Franka, and the Apptronik Apollo humanoid, by abstracting away robot-specific kinematic details into a shared generalized action space. This can be interpreted as a population-level abstraction at training time: MT learns a representation that is sufficient for motor control across the full distribution of embodiment configurations in the training fleet, discarding morphological variation that does not affect the underlying physics of the task. Specifically, MT operationalizes the information-bottleneck objective over the population of embodiments by minimizing $I(m_t; X)$, the compression of embodiment-specific history, while maximizing $I(m_t; Y)$, the predictive value for task-relevant motor behavior.

In $\pi_{0.7}$, the same principle is instantiated via multimodal conditioning: $\pi_{0.7}$ is trained with prompts that include not only task descriptions but episode metadata encoding how tasks were performed across the training population (for example, execution quality, speed, strategy) alongside control modality labels and visual subgoal images. Suboptimal trajectories are also incorporated by annotating with appropriate quality and speed metadata. Based on the analysis in Intelligence et al. [2026], the performance improvements observed with the inclusion of the conditioning information seem to indicate to us that the metadata might act as a sufficient statistic for the agent's behavioral class within the training population. Therefore, our hypothesis suggests that this metadata conditioning is a static, training-time mean field signal: a compressed summary of the population's behavioral distribution injected into each individual agent's policy as a conditioning input, instantiating the information-bottleneck objective over the offline training distribution rather than online fleet telemetry.

The two mechanisms are thus instantiations of the same underlying principle of encoding population-based distributions via mean-field learning (Section 3), which supports our position. MT abstracts over the heterogeneity of the training population, whereas diverse conditioning abstracts over the behavioral strategy dimension. Both demonstrate empirically that such population-level abstractions exist, are learnable from finite data, and substantially improve generalization, which is precisely the existence condition required for mean-field theory to guarantee tractable policy optimization. Taken together, they validate the mean-field framework not as a theoretical imposition on these systems but as the formal structure those systems are already approximating.

A.3 Compositional Task Generalization and Embodiments as Agentic Emergence

Both Team et al. [2025] and Intelligence et al. [2026] provide empirical evidence of agentic emergence, manifesting through distinct but structurally related phenomena: compositional task generalization in $\pi_{0.7}$, and cross-embodiment generalization in GR 1.5.

$\pi_{0.7}$ explicitly identifies compositional generalization, the ability to recombine skills learned from distinct training tasks to solve tasks never seen in training, as its central emergent capability. Intelligence et al. [2026] reports that $\pi_{0.7}$ can operate a kitchen air fryer having seen only two episodes of closing an air fryer and unrelated DROID dataset manipulation data; it can fold laundry on a bimanual UR5e platform having collected no laundry folding data for that robot, achieving success rates matching expert human teleoperators attempting the same zero-shot transfer. This emergent capability is threshold-like in the sense of Section 4. It is absent in task-specific fine-tuned specialists operating below the diversity threshold, and reliably present in $\pi_{0.7}$ operating above it, with the same per-agent policy architecture in both cases. The population measure $\mu_t^{(N)}$ in this context is the empirical distribution over training-time behavioral strategies and embodiment configurations, which is exactly the distribution that the diverse conditioning mechanism compresses into the metadata signal and that the mean field framework tracks dynamically at deployment.

GR 1.5's Motion Transfer mechanism leads to the same threshold-based emergence phenomenon observed over the morphology: a single VLA checkpoint achieves zero-shot control across ALOHA, Bi-arm Franka, and Apollo humanoid platforms by abstracting the population of embodiments into a shared action space, a capability that does not exist in platform-specific fine-tuned models and emerges only at sufficient embodiment diversity in the training population. Such deployment settings with heterogeneous platform types follow the frameworks of multi-class and graphon mean fields, and we believe could be avenues for future research in proposing a principled formalism of agentic emergence in robotics.