Exploring Spontaneous Social Interaction Swarm Robotics Powered by Large Language Models

Yitao Jiang[1], Luyang Zhao[1], Alberto Quattrini Li[1], Muhao Chen[2], and Devin Balkcom[1]
[1]Department of Computer Science, Dartmouth College, Hanover, NH, USA
Email: {yitao.jiang.gr, luyang.zhao.gr, alberto.quattrini.li, devin.balkcom}@dartmouth.edu
[2]Department of Aerospace Engineering, University of Kentucky, Lexington, KY, USA
Email: muhao.chen@uky.edu

***Abstract*— Traditional swarm robots rely on specific communication and planning strategies to coordinate particular tasks. Human swarms exhibit distinctive characteristics due to their capacity for language-based communication and active reasoning. This paper presents an exploratory approach to robotic swarm intelligence that leverages Large Language Models (LLMs) to emulate human-like active problem-solving behaviors. We introduce a decentralized multi-robot system where each robot initially only has its local information and does not know others' existence. The robots utilize LLMs for reasoning and natural language for inter-robot communication, enabling them to discover peers, share information, and coordinate actions dynamically. In a series of experiments in zero-shot settings, we observed human-like social behaviors, including mutual discovery, identification, information exchange, collaboration, negotiation, and error correction. While the technical approach is straightforward, the main contribution lies in exploring the interactive societies that LLM-driven robots form – a form of "robot anthropology" that examines emergent collaborative structures.**

***Index Terms*— Swarm Robotics, Swarm Intelligence, Large Language Model, Artificial Intelligence, Robot Anthropology, AI-Enabled Robotics, Multi-Robot Systems**

I. Introduction

In nature, ants collaborate to transport food and follow the trails of their predecessors [1]; fish schools collectively evade predators [2]; birds form specific formations during flight [3]; sheep exhibit synchronized movement patterns [4]; and wolves and hunting dogs demonstrate even more sophisticated patterns of collective intelligence during hunting [5]. Such inherent, passive swarm intelligence inherently limits complex information exchange [6], [7]. Human collective intelligence relies heavily on spoken and written communication. This capability for communication enables humans to collaborate on complex tasks that may not be predefined. [8], [9] Due to their inherent complexity, such tasks often exceed the capabilities of a single individual, necessitating collective effort [10]. Humans actively identify problems, analyze situations, organize groups, and collaborate to achieve solutions.

Large Language Models (LLMs) offer a transformative opportunity to enhance swarm robotics by endowing robots with more human-like social intelligence. Our architecture is shown in Fig. 1.

This paper explores a decentralized paradigm where robots, each driven by an independent LLM, spontaneously discover peers, establish communication through natural language, and self-organize collaborative behaviors without preprogrammed relationships. The primary contribution of this work is the investigation of emergent social dynamics in such LLM-driven swarms — a form of "robot anthropology." In experiments encompassing formation control and cooperative object transportation, we observed spontaneous peer discovery, dynamic negotiation, error correction, and other complex social interactions.

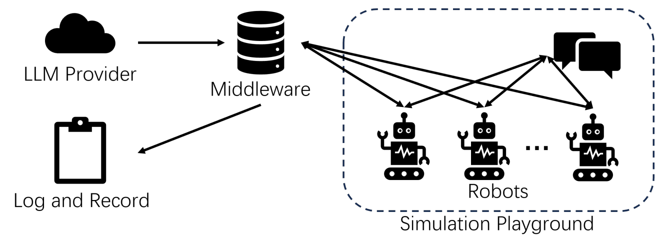

Fig. 1. System architecture overview showing the interaction between robots in the virtual environment, the proxy middleware that manages communication with LLM APIs, and the context management system. Each robot maintains an independent session that enables isolated reasoning and inter-robot communication.

Our key contributions include (1) a fundamentally different paradigm for swarm robot control devoid of preknowledge of peers and tasks, (2) experiments to analyze emergent social behaviors, and (3) a comprehensive analysis of task-adaptive communication patterns and baseline assessment of LLM capabilities in our application.

These findings demonstrate the feasibility of human-inspired active swarm intelligence and represent a step towards more adaptable, language-driven robotic systems capable of emergent, cooperative problem-solving.

II. Related Work

A. Traditional Swarm Robotics Approaches

In the field of swarm robotics, traditional approaches, including formation control [11], flocking algorithms [12], consensus-based approaches [13], and bio-inspired swarm algorithms [14], have shown effectiveness in predictable environments. However, relying on predefined behavioral patterns, these methods often lack adaptability and flexibility when facing open-ended tasks [15], [16]. Zhou and

Tokekar examined multi-robot coordination in uncertain environments, focusing on algorithmic planning approaches for adaptive decision-making, yet still within structured frameworks [17]. Similarly, Gielis et al. provided a critical analysis of communication mechanisms in multi-robot systems, emphasizing the need for efficient information exchange protocols while highlighting the limitations of conventional methods [18]. Building on these challenges, Korsah et al. developed a comprehensive taxonomy for multi-robot task allocation that maps robotic challenges to established mathematical optimization models, offering systematic classification but still within traditional paradigms [19]. In the classic paradigms, [20] and [21] highly rely on human control, while some automation algorithms appeared in the inter-robot collaboration in [22] and [23], but still unable to self drive to accomplish the tasks.

B. AI-Driven Agents

Our approach differs from recent LLM-based game agents. While frameworks like ALYMPICS [24], LLM agent societies in Avalon [25], LARP for role-playing [26], and other game agents across various genres [27] demonstrate impressive strategic decision-making and social behaviors, they operate within predefined rules and structured scenarios. In contrast, our system creates an open-ended environment where robots organically develop collaboration strategies that demonstrate deliberate communication and logical deduction that more closely resembles human problem-solving.

C. LLM Applications in Robotics

Recent breakthroughs in LLMs have opened new possibilities in robot control. LLM2Swarm pioneered the integration of LLMs into robot swarms through two approaches: indirect integration for controller synthesis and validation and direct integration deploying local LLM instances on each robot for collaboration and human-robot interaction [28]. While this work demonstrated LLM's potential in reasoning, planning, and collaboration, it primarily utilized LLMs as task planners and controllers within predetermined collaboration patterns. Li et al. systematically compared different LLM-based communication frameworks (DMAS, CMAS, HMAS-1, HMAS-2) in multi-robot systems, focusing on system scalability and task success rates [29]. Lykov and Tsetserukou developed LLM-BRAIn, a transformer-based LLM fine-tuned to generate adaptive robot behaviors via behavior trees (BTs), trained on 8.5k GPT-3.5 demonstrations and performs comparably to human-created BTs [30]. Liu et al. proposed a Human-Robot Collaboration (HRC) approach using GPT-4 and YOLO-based perception to enhance LLM-based robotics, enabling complex task execution through human-guided learning and motion planning [31]. Wang et al. addressed LLMs' limitations in embodied robot tasks by proposing a multimodal GPT-4V framework that integrates language and visual inputs, enhancing robot performance and advancing Human-Robot-Environment interaction [32].

Table I highlights key differences between our approach and representative works. Bio-inspired methods [12], [14] typically rely on predetermined relationships with limited communication, while recent LLM-based approaches [28], [29] introduce language capabilities but generally within structured interaction frameworks. In contrast, our approach enables spontaneous social interaction, where robots initially have no knowledge of others' existence and must actively discover peers, establish communication, and self-organize.

TABLE I
COMPARISON WITH REPRESENTATIVE WORKS

Feature	**Bio-inspired** [12], [14]	**LLM-based** [28], [29]	**Our Approach**
Robot Discovery	Typically predefined	Often predetermined	Spontaneous
Language Use	Minimal/ Symbolic	Task-specific	Open-ended
Task Adaptation	Fixed algorithms	Requires specific prompts	Generic prompts
Social Dynamics	Rule-based	Structured	Emergent

Note: Evaluations reflect trends in cited works and may not represent all implementations.

III. PROBLEM FORMULATION

A. System Design and Implementation

Our system is implemented in a simple virtual environment written in Python with OpenCV visualization. To isolate and study the phenomena of language-based social coordination, which is our primary research focus, we deliberately simplified physical properties such as collision detection. We developed a proxy middleware to unify the management of all communications with various LLM APIs, handle context management, and perform logging. This proxy middleware does not change the distributed nature of the agent decision-making system.

The proxy processes: (a) sending prompts or conversations from robots to the LLM; (b) receiving generated responses from the LLM and parsing to robot commands; and (c) managing context for each robot session to record logs, as shown in Fig. 1. Using this middleware rather than integrating these functions into the robot simulator reduces complexity and decouples the code while making it convenient to switch between different AI models.

When a robot connects to the proxy, the proxy creates an independent session for that robot. Each robot in the virtual environment maintains its independent context, isolated from other robots. All communication and context operations for a robot occur within its corresponding session, and the proxy records the context in real-time to files associated with that session. Robots and humans can broadcast messages within the virtual environment, which are received by other robots and processed by their respective LLMs, enabling inter-robot communication.

B. Experimental Design

We demonstrate the spontaneous communication and collaboration capabilities of our system through eight tasks, as shown in Fig. 2, which can be classified into two categories. For Tasks 1-5, we focus on exploring formation control and geometric reasoning, where robots must communicate, exchange positional information, and reason about spatial relationships to achieve structured formations, such as alignment, triangles, and circles. Tasks 6-8 focus on cooperative object transportation, where robots must coordinate their roles, negotiate task allocation, and execute clever and assistive transportation of objects. Specifically, Task 8 highlights sequential task execution and coordination, where robots must relay objects within a constrained movement range, demonstrating adaptive teamwork and stepwise collaboration.

- **Task 1 – Mutual Face-to-face Alignment:** Two randomly placed robots must face each other, requiring them to discover each other's presence, inquire about positions, and reason about necessary rotations.
- **Task 2 – Robots Alignment:** Four robots randomly placed along the y (evenly distributed on the x for better visualization in the experiments) must align on the same y-value, demonstrating multi-agent discovery, position information exchange, goal position negotiation, and task completion verification.
- **Task 3 – Equilateral Triangle Formation:** Three robots must form an equilateral triangle, testing geometric reasoning capabilities.
- **Task 4 – More Complex Triangle Formation:** Four robots must organize into a triangle formation, with one robot necessarily positioned along an edge, testing autonomous coordination when perfect symmetry is impossible.
- **Task 5 – Circle Formation:** Six robots must form a circle, challenging the task with more robots involved than any of the other tasks.
- **Task 6 – Single Object Transportation:** Two robots and one object are placed in the environment, with the task of moving the object (which requires only one robot to transport) to a target location. Thus, this tests efficient task allocation when only one robot needs to complete the task.
- **Task 7 – Dual Object Transportation:** Similar to Task 6, but with two objects instead of one, thus increasing the number of possible robot-object pairings.
- **Task 8 – Relay Transportation:** Three robots with restricted movement ranges must coordinate to transport one object, which can be carried by one robot at a time, to a target location. Because of the range restriction, the robots must relay the objects to one another.

We expect traditional swarm robotics methods would require separate algorithms for each demo, with additional coding needed for more complex requirements. Our LLM-based approach represents a more adaptive solution without task-specific programming.

C. LLM Setup

We mainly experimented with GPT-4o-2024-11-20, which provided the best results on our preliminary experiments, but we also tested how other readily available LLMs perform to evaluate generalization to other LLMs. The tests share the same prompt, and the temperature is set to 0.7. For standardized control command output, we employed GPT-4o-mini as a formatting tool, using its JSON output capabilities but without using it for any of the reasoning tasks.

TABLE II
SUCCESS COUNTS (OUT OF 10 TRIALS) FOR LLMS ON EXPERIMENTAL TASKS.

Model	T1	T2	T3	T4	T5	T6	T7	T8
GPT-4o	6	8	6	4	1	7	5	1
Gemini-2.0-Flash	5	8	5	0	0	7	3	1
DeepSeek-V3	4	1	2	0	0	2	1	0

We ran 10 trials on GPT-4o-2024-11-20, DeepSeek-V3, and Gemini-2.0-Flash-001 for each experiment. We define failure conditions as when the robots cease to generate any new communication interactions or fail to correctly complete the task within a specified time window. Tasks 1-7 have 10 minutes of timeout, while Task 8 has 15 minutes of timeout, given its additional complexity. The success attempts were recorded in Table II. The detailed logs, recordings, and data can be accessed in the supplementary materials.

IV. OBSERVATION AND CHALLENGES

A. Results and Analysis

As shown in Table II, GPT-4o achieved the highest success rates across all eight tasks, with a particularly strong performance in Task 2 (Robot Alignment, 8/10) and Task 6 (Single Object Transportation, 7/10). Gemini-2.0-Flash demonstrated comparable results on simpler tasks but struggled with more complex geometric reasoning in Tasks 4 and 5. DeepSeek-V3 showed significantly lower success rates across all experiments, even when given 5x time limits.

Beyond the three main LLMs in our experiment, we conducted limited tests with several other models. Grok-2 and Claude-3.7-Sonnet successfully completed at least a few tasks, but API request limitations prevented comprehensive testing across all experimental scenarios. Claude-3.5-Sonnet exhibited severe hallucination tendencies, frequently generating irrelevant messages, inventing non-existent information, or prematurely declaring successful completion of tasks. GPT-4o-mini demonstrated extremely limited context retention, often forgetting critical information after just 3-4 exchanges and also frequently generating repeated meaningless text. These observations indicate that effective multi-robot coordination through natural language using the approach we took may require substantial reasoning capacity and context management capabilities that appear to be available only in larger, more advanced LLMs.

Analysis of failure cases revealed several common patterns. In unsuccessful trials, robots frequently misinterpreted their objectives or made critical errors in mathematical calculations when determining formation coordinates. For example, DeepSeek-V3 always misunderstood the requirement of uniform distribution in Task 5, so that robots reach the circle but are not spread evenly. All LLMs frequently calculate the orientation wrong, which causes them not to head to the

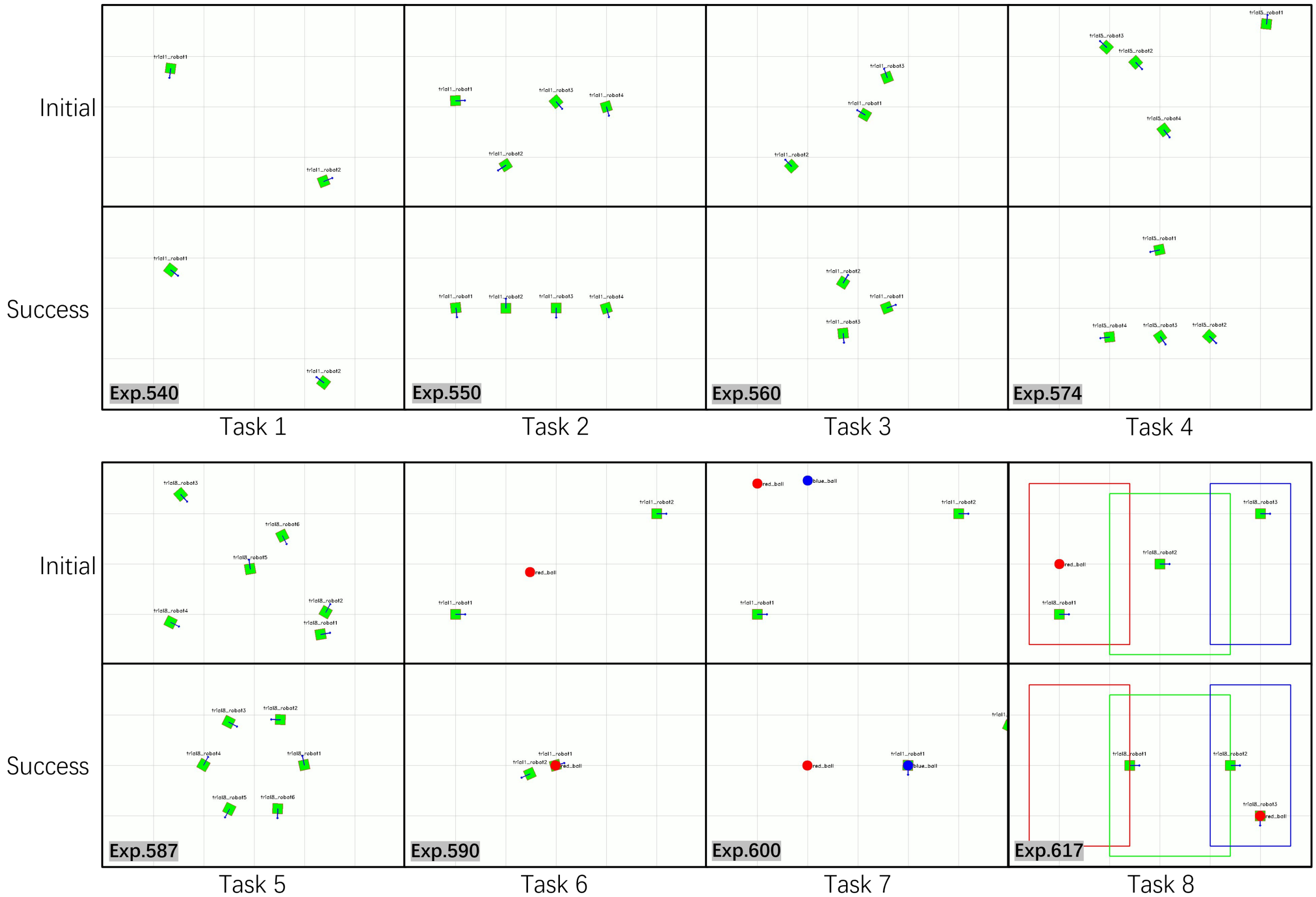

Fig. 2. Illustration of the eight experimental scenarios: Tasks 1-5 explore formation control and geometric reasoning (mutual alignment, robot alignment, equilateral triangle, complex triangle, and circle formations), while Tasks 6-8 demonstrate cooperative object transportation (single object, dual object, and relay transportation).

target destination, but this type of error is recoverable. All LLMs may also generate commands that do not follow the rules stated in the system prompt, which causes silence (i.e., no communication exchange) between robots and consequent inaction, leading to eventual failure.

Tasks requiring precise geometric reasoning with multiple agents (Tasks 4, 5) or sequential coordination (Task 8) proved the most challenging. The circle formation task (Task 5) was particularly difficult, with only one successful completion using GPT-4o. This suggests that as the number of robots increases, the dimensional complexity of spatial reasoning and communication grows non-linearly, exceeding the current capabilities of most LLMs.

B. Communication Pattern Analysis

Analysis of communication patterns in successful task executions reveals distinct interaction strategies across different task types.

Tasks 1 to 8 required an average of 8, 11.75, 5.67, 8, 7.5, 11.5, 15.5, and 14.33 communications to complete.

We classified robot communications into eight categories: Status Report (reporting current position, status, or progress), Query (requesting information or confirmation), Plan Announcement (declaring intentions or plans), Coordination (organizing or directing other robots), Help Request (explicitly asking for assistance), Help Offer (providing help or solutions), Acknowledgment (confirming information or task completion), and Other (human instruction or communications not fitting previous categories). We utilized GPT-4o to label each message in all the conversations. As shown in Fig. 3, while our system prompts note they can collaborate, the prompt does not indicate that the robots should use any dictated communication structures or patterns.

Status Reports dominated across all tasks (38-56% of messages), with robots regularly sharing position and state information. The highest proportion appeared in Task 2 (56%), where accurate alignment requirements apparently led to frequent position updates. The formation tasks generally showed higher rates of Status Reports compared to transportation tasks, reflecting the continuous positional adjustments needed for geometric arrangements.

Task-specific communication patterns emerged clearly in our data. Formation tasks (1-5) showed minimal Help Requests (0%) but substantial Acknowledgments (up to 30% in Task 5), indicating a coordination-focused approach where consensus building was critical. In contrast, transportation

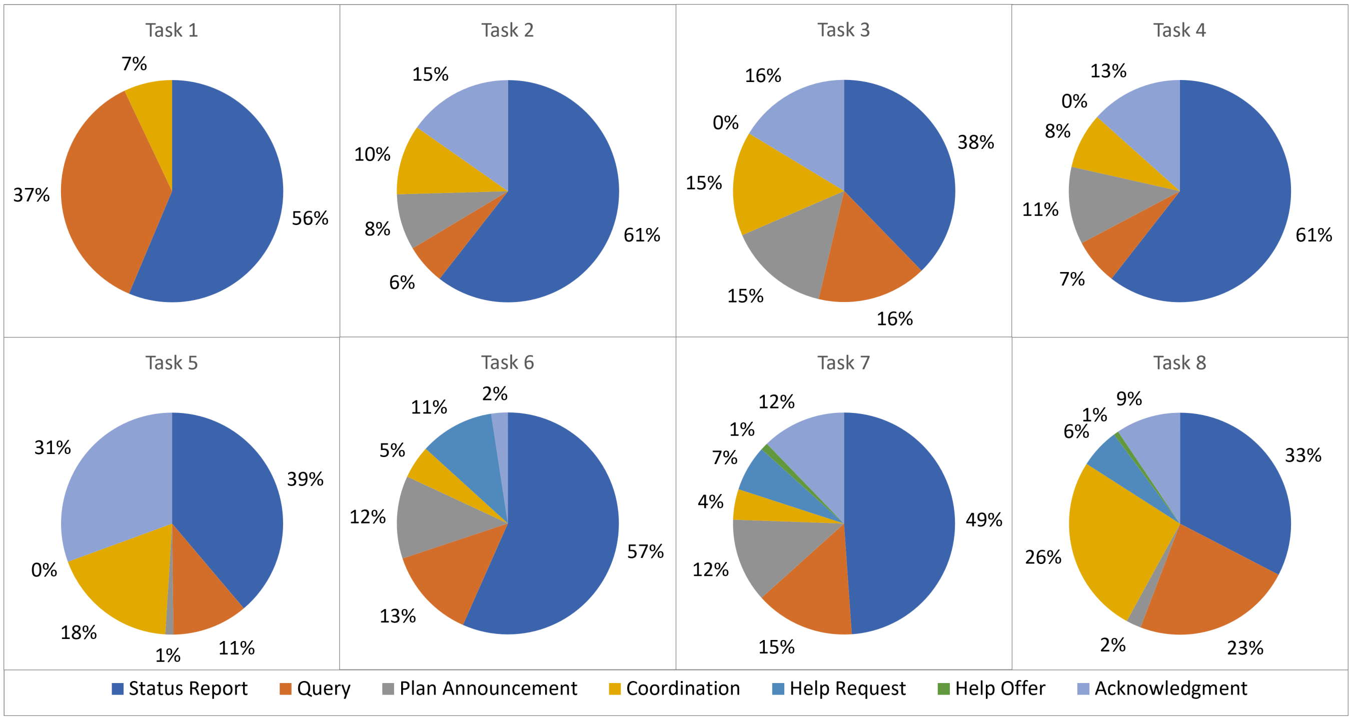

Fig. 3. Distribution of communication message types across the eight experimental tasks successes with GPT-4o-2024-11-20. The chart shows how communication patterns appear in relation to the task requirements.

tasks (6-8) exhibited more Help Requests (6-9%) and reduced Acknowledgments (2-11%), perhaps reflecting a more direct problem-solving approach when physical manipulation was required.

Query messages showed task-dependent patterns, with the highest proportions in Task 1 (30%) and Task 8 (23%). This reflects the information-gathering requirements of these specific scenarios – mutual discovery in Task 1 and complex relay coordination in Task 8. The high proportion of Coordination messages in Task 8 (26%) further demonstrates how communication adapts to sequential dependency requirements.

Examining the ratio between information sharing (Status Reports + Queries) and coordination messages (Plan Announcements + Coordination) reveals a task-dependent evolution:

- Simple discovery (Task 1): 92.9% vs. 7.0%
- Intermediate formation (Tasks 2-4): ∼66% vs. ∼19%
- Complex formation (Task 5): 49.8% vs. 19.6%
- Transportation tasks (6-7): ∼66% vs. ∼17%
- Sequential transportation (Task 8): 55.8% vs. 28.3%

This progression shows how robots naturally shift from information-heavy to coordination-heavy communication as task complexity increases, particularly when sequential dependencies are involved. Task 8 (relay transportation) exhibited both high Coordination (26.1%) and Query (23.2%) rates, directly reflecting the sequential dependencies required in relay operations. These proportions significantly exceed those in simpler tasks, demonstrating how communication naturally adapts to coordination complexity.

Formation tasks (1-5) and transportation tasks (6-8) exhibited substantially different communication distributions. Most notably, Help Requests and Help Offers (combined 0% in formation tasks) emerged as significant components in transportation tasks (7-11%), reflecting the physical interdependencies inherent in manipulation tasks. Task 1 showed the highest proportion of Queries (36.6% of meaningful messages) and no Acknowledgments (0%), revealing a discovery-focused communication strategy. In contrast, Task 5 (circle formation) showed the highest proportion of Acknowledgments (30.6%), reflecting the increased need for confirmation in complex spatial arrangements.

C. Emergent Social Behaviors

In our experiments with LLM-driven robot swarms, we observed several social behaviors that emerged naturally through multi-agent interactions. While LLMs inherently possess conversational abilities, these observed behaviors manifest uniquely in multi-robot environments and cannot exist in single-agent scenarios. The behaviors we describe below emerge from the robots' ability to reason about other robots' states, intentions, and needs. This capability fundamentally distinguishes our approach from both traditional swarm robotics methods and single-agent LLM applications.

1) Collaborative Mathematical Optimization: In multiple trials, robots autonomously performed mathematical reasoning to optimize group behavior. For example, in one trial of Task 2, shown in Fig. 4(a), when tasked with aligning to a common y-value, the robots shared their positions and calculated the optimal alignment target.

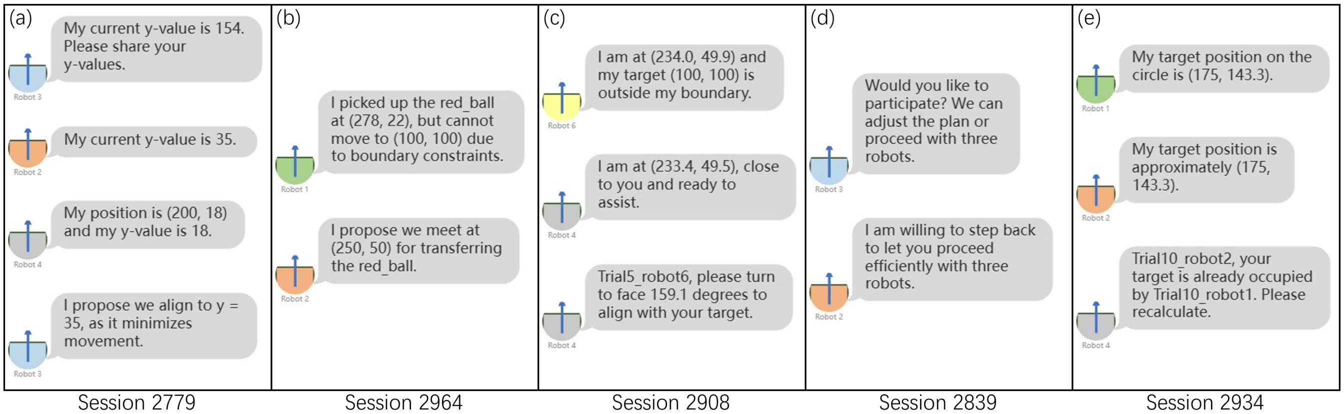

Fig. 4. Showcase of emergent social behaviors, extract from a trial in Tasks 2, 8, 6, 4, and 7. (sessions 2779, 2964, 2908, 2839, and 2934).

2) Adaptive Resource-Constrained Coordination: When faced with boundary constraints that prevented direct task completion, robots spontaneously devised handoff strategies. In Session 2964, shown in Fig. 4(b), a robot recognized its inability to complete the task alone.

This coordination emerged without pre-programmed handoff protocols, demonstrating the robots' ability to decompose problems based on individual constraints, a capability not typically seen in traditional swarm systems.

3) Personalized Assistance Behaviors: We observed instances where robots provided detailed guidance to help others overcome difficulties. In Session 2908, shown in Fig. 4(c), when one robot encountered boundary constraints.

This teaching-like behavior demonstrates knowledge sharing and assistance not typically observed in traditional swarm approaches.

4) Team Efficiency Meta-Reasoning: In several trials, robots demonstrated meta-reasoning about optimal team composition. Session 2839, shown in Fig. 4(d), provides an example where a robot voluntarily removed itself.

This self-reflective optimization represents a sophisticated social awareness absent in traditional swarm approaches, which typically utilize all available units regardless of optimal team size.

5) Predictive Conflict Management: Robots demonstrated the ability to detect and resolve potential conflicts before they occurred. In Session 2934, shown in Fig. 4(e), when two robots targeted the same position.

This proactive conflict detection based on awareness of others' declared intentions rather than physical collisions demonstrates predictive social coordination that extends beyond reactive collision avoidance typically employed in traditional swarm robotics.

D. Critical Failure Modes

We selectively choose to analyze several significant failure modes specific to our LLM-driven robot swarms. We can learn from it and figure out the cause of failure. These patterns reveal fundamental research challenges at the intersection of language models and multi-robot systems.

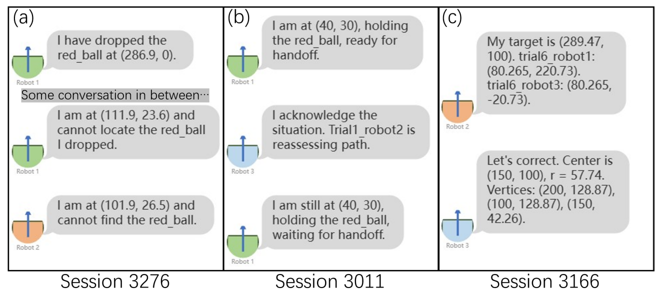

Fig. 5. Showcase of critical failure modes, extract from Tasks 1, 1, and 6 (sessions 3276, 3011, and 3166).

1) Object State Tracking Inconsistency: LLM-driven robots demonstrated difficulty maintaining consistent object tracking after interactions. In Session 3276, shown in Fig. 5(a), robots lost track of an object after dropping it.

Unlike traditional robotic systems with explicit object state representations, LLM-driven systems rely on natural language state updates, which are vulnerable to information loss during extended interactions.

2) Communication Loop Entrapment: In several trials, robots became trapped in circular communication patterns without task progress. Session 3011, shown in Fig. 5(b), demonstrates this phenomenon.

This pattern persisted for dozens of exchanges without progress. The social communication patterns generated by LLMs, while impressively human-like, can lead to inefficient coordination compared to more direct protocols used in traditional approaches.

3) Geometric Reasoning Failures: LLM-driven robots frequently exhibited significant errors in spatial reasoning and geometric calculations. In Session 3116, shown in Fig. 5(c), multiple robots calculated incorrect positions for an equilateral triangle.

Despite multiple correction attempts, the proposed coordinates remained mathematically invalid. This reveals a fundamental limitation in LLMs' ability to perform consistent mathematical calculations, which is a capability essential for successful swarm robotics operations.

These failure modes highlight important areas for improvement in LLM-based swarm control. Future implementations should focus on enhancing world-state modeling consistency, developing structured communication protocols to prevent circular patterns, and incorporating validation mechanisms for mathematical calculations.

E. Response Time and Performance Analysis

Beyond task execution failures themselves, a significant challenge in our experiments involved LLM API reliability. In a multi-robot environment, API requests that are excessively delayed or failed cannot simply be retried, as the interaction context evolves continuously. Despite implementing delay mechanisms and timeout parameters to mitigate these issues, they remained a notable concern throughout our experiments. We observed that certain task failures stemmed not from inherent LLM reasoning limitations but from API instability.

Identifying and isolating these API-related failures in batch experiments is challenging. We chose not to manually filter such failures from our results, as doing so would potentially introduce bias and reduce result fidelity. As our primary objective was to establish a proof of concept, successfully executed tasks sufficiently demonstrated the viability of our approach, with failure cases and success rates providing supplementary insights.

DeepSeek-V3 exhibited particularly pronounced latency and API instability during our experiments, with response times ranging from 5 seconds to several minutes, occasionally returning empty responses. As previously noted, we implemented extended timeout parameters 5 times for DeepSeek-V3, but this intervention produced no measurable improvement in performance outcomes. Task failures resulting from communication silence typically occurred well before timeout thresholds were reached.

Although we provided a comprehensive communication pattern analysis, we did not report a detailed analysis of LLM response latency metrics, as they depend on a number of factors beyond our experimental control, including Internet connection speed and service queue, and they don't affect the fundamental contribution of this work: the novel paradigm of LLM-enabled swarm robots spontaneously discovering peers, self-organizing, and coordinating task execution through language-based interaction. We expect that in the case of actual deployments for multi-robot systems, there would be a dedicated service with a reliable connection.

F. Discussion and Limitations

Despite interesting results, our approach faces several significant limitations that suggest future work.

First, LLMs exhibit fundamental weaknesses in consistent mathematical reasoning, particularly evident in geometric formation tasks where calculation errors frequently lead to task failures. The computational demands and API response latency issues also present practical challenges for real-time robotic applications. In our experiments, we observed that API reliability varied considerably between models, with some requests experiencing delays ranging from seconds to minutes or failing entirely. These technical limitations significantly impacted experimental outcomes but were deliberately not filtered from our results to maintain data integrity.

Achieving high success rates is not our primary objective right now. We find there is a large potential to optimize for task completion. We expect a few-shot strategy and fine-tuning to be our next approach. We will also figure out the engineering solution to handle unreliable API calls.

Our current simulation environment also made several simplifying assumptions, particularly in omitting collision detection, sensor limitations, and physical constraints. While useful for initial proof of concept, these simplifications do not fully represent real-world robotic challenges. Future work must address physical implementation concerns, including sensor noise, limited perception, unreliable communication channels, and physical interaction constraints.

We will also build real robots and experiment in the real world to investigate inter-robot collaboration further. It would be even more interesting if we made different heterogeneous robots that have different functionalities. We can further investigate how different functional robots collaborate spontaneously.

V. Conclusion

Our research expands the conceptual boundaries of swarm robotics by integrating human-like social intelligence capabilities through LLMs. While traditional swarm approaches excel at specific tasks through pre-programmed behavioral patterns, they lack generalized problem-solving abilities. Our decentralized approach, where each robot maintains independent reasoning without central control, preserves core swarm principles while adding dimensions of adaptability through natural language reasoning. The demonstrated ability of robots to discover peers, establish communication, and self-organize for diverse tasks without task-specific programming represents a qualitative advance in swarm flexibility and autonomy.

The most significant finding from our experiments is the emergence of sophisticated social behaviors that resemble human collaborative patterns. These include collaborative mathematical optimization, where robots collectively reasoned about optimal positioning; adaptive resource-constrained coordination, where robots devised handoff strategies based on individual limitations; personalized assistance behaviors, including teaching-like guidance to peers; team efficiency meta-reasoning with voluntary role adjustments; and predictive conflict management through intention-based coordination.

Our communication pattern analysis revealed task-specific adaptations in robot dialogue, with proportions of status reports, queries, and coordination messages naturally shifting based on task requirements. As task complexity increased, particularly in scenarios with sequential dependencies, robots naturally evolved more coordination-heavy communication strategies. The stark differences between communication patterns in formation tasks versus transportation tasks further demonstrate how LLM-driven robots can adapt their interaction styles to task demands without explicit programming towards generalized swarm robotics.

Supplementary Materials

All data are open-sourced on GitHub: `https://github.com/cccat6/LLM-Swarm`.

Acknowledgment

We have used generative AI, including GPT-4o and Claude-3.7-Sonnet, to extract and summarize data, implement code, and help polish the paper writing.

References

[1] H. F. McCreery and M. Breed, "Cooperative transport in ants: a review of proximate mechanisms," *Insectes sociaux*, vol. 61, pp. 99–110, 2014.

[2] U. Lopez, J. Gautrais, I. D. Couzin, and G. Theraulaz, "From behavioural analyses to models of collective motion in fish schools," *Interface focus*, vol. 2, no. 6, pp. 693–707, 2012.

[3] I. L. Bajec and F. H. Heppner, "Organized flight in birds," *Animal Behaviour*, vol. 78, no. 4, pp. 777–789, 2009.

[4] J. Gautrais, P. Michelena, A. Sibbald, R. Bon, and J.-L. Deneubourg, "Allelomimetic synchronization in merino sheep," *Animal Behaviour*, vol. 74, no. 5, pp. 1443–1454, 2007.

[5] A. Berghänel, M. Lazzaroni, G. Cimarelli, S. Marshall-Pescini, and F. Range, "Cooperation and cognition in wild canids," *Current Opinion in Behavioral Sciences*, vol. 46, p. 101173, 2022.

[6] R. C. Eberhart, Y. Shi, and J. Kennedy, *Swarm intelligence*. Elsevier, 2001.

[7] Y. Wang, H. Chen, L. Xie, J. Liu, L. Zhang, and J. Yu, "Swarm autonomy: From agent functionalization to machine intelligence," *Advanced Materials*, vol. 37, no. 2, p. 2312956, 2025.

[8] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm intelligence: from natural to artificial systems*. Oxford University Press, 1999, no. 1.

[9] T. Kameda, W. Toyokawa, and R. S. Tindale, "Information aggregation and collective intelligence beyond the wisdom of crowds," *Nature Reviews Psychology*, vol. 1, no. 6, pp. 345–357, 2022.

[10] K. H. Petersen, N. Napp, R. Stuart-Smith, D. Rus, and M. Kovac, "A review of collective robotic construction," *Science Robotics*, vol. 4, no. 28, p. eaau8479, 2019.

[11] Y. Zhang, S. Oğuz, S. Wang, E. Garone, X. Wang, M. Dorigo, and M. K. Heinrich, "Self-reconfigurable hierarchical frameworks for formation control of robot swarms," *IEEE transactions on cybernetics*, vol. 54, no. 1, pp. 87–100, 2023.

[12] L. E. Beaver and A. A. Malikopoulos, "An overview on optimal flocking," *Annual Reviews in Control*, vol. 51, pp. 88–99, 2021.

[13] R. Yan and A. Julius, "Distributed consensus-based online monitoring of robot swarms with temporal logic specifications," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9413–9420, 2022.

[14] S. Tianbo, H. Weijun, C. Jiangfeng, L. Weijia, Y. Quan, and H. Kun, "Bio-inspired swarm intelligence: a flocking project with group object recognition," in *International Conference on Consumer Electronics and Computer Engineering (ICCECE)*. IEEE, 2023, pp. 834–837.

[15] J. Kuckling, "Recent trends in robot learning and evolution for swarm robotics," *Frontiers in Robotics and AI*, vol. 10, p. 1134841, 2023.

[16] D. Garzón Ramos, F. Pagnozzi, T. Stützle, and M. Birattari, "Automatic design of robot swarms under concurrent design criteria: A study based on iterated f-race," *Advanced Intelligent Systems*, vol. 7, no. 1, p. 2400332, 2025.

[17] L. Zhou and P. Tokekar, "Multi-robot coordination and planning in uncertain and adversarial environments," *Current Robotics Reports*, vol. 2, pp. 147–157, 2021.

[18] J. Gielis, A. Shankar, and A. Prorok, "A critical review of communications in multi-robot systems," *Current Robotics Reports*, vol. 3, pp. 213–225, 2022.

[19] G. A. Korsah, A. Stentz, and M. B. Dias, "A comprehensive taxonomy for multi-robot task allocation," *The International Journal of Robotics Research*, vol. 32, no. 12, 2013.

[20] L. Zhao, Y. Wu, J. Blanchet, M. Perroni-Scharf, X. Huang, J. Booth, R. Kramer-Bottiglio, and D. Balkcom, "Soft lattice modules that behave independently and collectively," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 1–1, July 2022.

[21] L. Zhao, Y. Wu, W. Yan, W. Zhan, X. Huang, J. Booth, A. Mehta, K. Bekris, R. Kramer-Bottiglio, and D. Balkcom, "Starblocks: Soft actuated self-connecting blocks for building deformable lattice structures," *IEEE Robotics and Automation Letters*, vol. PP, no. 99, pp. 1–8, August 2023.

[22] L. Zhao, Y. Jiang, B. Kostas, M. Chen, and D. Balkcom, "Tensegrity-blocks: Modular shape-changing blocks enable self-assembling robotic structures," February 2024.

[23] L. Zhao, Y. Jiang, C.-Y. She, Q. L. Alberto, C. Muhao, and D. Balkcom, "Softrafts: Floating and adaptive soft modular robots," December 2024.

[24] S. Mao, Y. Cai, Y. Xia, W. Wu, X. Wang, F. Wang, T. Ge, and F. Wei, "ALYMPICS: LLM agents meet game theory – exploring strategic decision-making with AI agents," 2024.

[25] Y. Lan, Z. Hu, L. Wang, Y. Wang, D. Ye, P. Zhao, E.-P. Lim, H. Xiong, and H. Wang, "LLM-based agent society investigation: Collaboration and confrontation in avalon gameplay," 2024.

[26] M. Yan, R. Li, H. Zhang, H. Wang, Z. Yang, and J. Yan, "LARP: Language-agent role play for open-world games," 2023.

[27] S. Hu, T. Huang, F. Ilhan, S. Tekin, G. Liu, R. Kompella, and L. Liu, "A survey on large language model-based game agents," 2024.

[28] V. Strobel, M. Dorigo, and M. Fritz, "LLM2Swarm: robot swarms that responsively reason, plan, and collaborate through LLMs," *arXiv preprint arXiv:2410.11387, accepted at NeurIPS 2024 Workshop on Open-World Agents*, 2024.

[29] P. Li, Z. An, S. Abrar, and L. Zhou, "Large language models for multi-robot systems: A survey," *arXiv preprint arXiv:2502.03814*, 2025.

[30] A. Lykov and D. Tsetserukou, "LLM-brain: AI-driven fast generation of robot behaviour tree based on large language model," in *2024 2nd International Conference on Foundation and Large Language Models (FLLM)*. IEEE, 2024, pp. 392–397.

[31] H. Liu, Y. Zhu, K. Kato, A. Tsukahara, I. Kondo, T. Aoyama, and Y. Hasegawa, "Enhancing the LLM-based robot manipulation through human-robot collaboration," *IEEE Robotics and Automation Letters*, 2024.

[32] J. Wang, E. Shi, H. Hu, C. Ma, Y. Liu, X. Wang, Y. Yao, X. Liu, B. Ge, and S. Zhang, "Large language models for robotics: Opportunities, challenges, and perspectives," *Journal of Automation and Intelligence*, 2024.