

ChatAgents: Enhancing Autonomous Decision-Making in Multi-Agent Conversations via Intrinsic Motivation

1st Tao Huang^{1,2}

¹Hangzhou Institute of Medicine,
Chinese Academy of Sciences
Hangzhou, China

²University of Chinese Academy of Sciences
Beijing, China
hta36@163.com

2nd Junwei Zhang*

Hangzhou Institute of Medicine,
Chinese Academy of Sciences
Hangzhou, China

zhangjunwei@him.cas.cn

3rd Pengju Yan

Hangzhou Institute of Medicine,
Chinese Academy of Sciences
Hangzhou, China

yanpengju@gmail.com

5th Xiaolin Li^{1,2,3*}

¹Hangzhou Institute of Medicine,
Chinese Academy of Sciences
Hangzhou, China

²University of Chinese Academy of Sciences
Beijing, China

³School of Statistics and Data Science,
School of Information and Electronic Engineering,
Zhejiang AIX College, Zhejiang Gongshang University
Hangzhou, China
xiaolinli@ieee.org

4th Rui Yang

School of Nursing
Zhejiang Chinese Medical University
Hangzhou, China
kelseyjhonna@163.com

Abstract—Large Language Models (LLMs) exhibit advanced language capabilities, yet current agents lack genuine behavioral autonomy, operating on predefined logic rather than dynamically deciding when and how to participate in conversations. This paper proposes ChatAgents, a framework that endows agents with intrinsic motivation for autonomous decision-making in multi-agent group discussions. The framework introduces two key mechanisms: a context-driven module activator enabling dynamic memory retrieval and goal analysis based on real-time dialogue, and a preemptive speaking mechanism where agents autonomously compete for turn-taking based on internal willingness and simulated reaction times. In simulated team-building scenarios, agents exhibited emergent social behaviors—including spontaneous leadership, conflict mediation, and innovative proposal generation—without external intervention, while maintaining consistent personality traits throughout interactions. This work demonstrates a pathway toward more authentic autonomous systems by enabling LLM agents to replicate the complexity of human conversational dynamics.

Index Terms—Agent, Autonomous, Large Language Models, Decision-Making.

I. INTRODUCTION

The development of Large Language Models (LLMs)—exemplified by GPT [1], [2] and LLaMA [3]—

This work is supported in part by funds from the National Key Research and Development Program of China (2022YFC3600902), Key Research and Development Program of Zhejiang (2025C01129) and the National Science Technology Major Project (No. 2023ZD0509706).

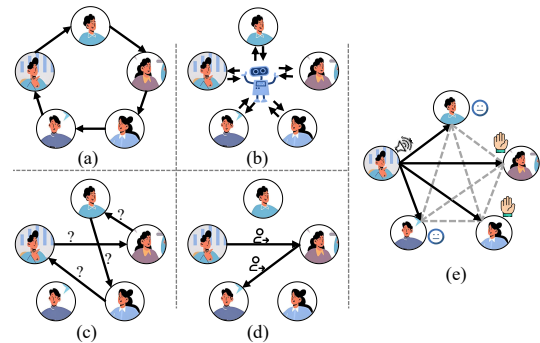


Fig. 1. Communication Orders in Multi-Agent Systems. (a) Rule-Based Order: Agents speak in a predefined sequence. (b) Centralized Order: A central controller manages speaking turns. (c) Random Order: Agents speak in an unstructured, probabilistic manner. (d) Designated Order: The current speaker chooses the next speaker. (e) Self-Driven Order: Agents decide autonomously when to speak based on internal state.

[5]—has greatly propelled progress in natural language understanding and generation [6]. This advancement has revived interest in AI agents—systems designed to autonomously perceive, decide, and act [7]. However, research on LLM-based agents has largely followed two paradigms: simulation-oriented (e.g., role-playing [8]) and task-oriented. Both approaches depend on predefined triggers or static instructions, thereby lacking dynamic autonomy

and limiting their applicability to open-ended interactions and the simulation of complex social behaviour [9], [10]. Although autonomy—as a core feature enabling independent decision-making—has been widely studied in multi-agent systems [11], its role in managing turn-order decisions within group chat environments remains under-explored. Existing systems typically rely on externally imposed, rule-driven mechanisms [12], which lack contextual adaptability and suppress proactive decision-making. As a result, they fail to reproduce human-like conversational dynamics, in which participants actively compete for speaking turns based on real-time intentions and goals.

To overcome the limitations of traditional rule-based, centralized, and random interaction models (Figure 1), we introduce the Self-Driven Order model. This paradigm enhances agent autonomy by replacing rigid turn-taking with dynamic, context-aware decision-making. In this framework, agents independently assess the relevance and timing of their contributions based on real-time contextual factors, such as participant states and specific goals. This design provides a more authentic representation of the spontaneous dynamics inherent in human conversation.

Drawing inspiration from the human mind’s capacity for dynamic cognitive coordination, this paper presents ChatAgents, a multi-agent cognitive architecture driven by dynamic prompting. Designed to facilitate human-like, context-aware autonomy, the framework integrates two primary mechanisms: context-driven module activation and preemptive speech decision-making. The former enables agents to autonomously invoke specific cognitive functions—such as memory retrieval and goal inference—based on real-time conversational cues and internal objectives. The latter empowers agents to independently initiate turn-taking by evaluating utterance urgency and informational utility, ensuring natural and coherent multi-turn interactions.

We evaluated ChatAgents’s capacity for human-like autonomy through simulated corporate team-building discussions scenarios. Operating without external guidance, the agents generated role-consistent dialogue based solely on initial configurations. Our analysis focused on emergent social behaviors—including spontaneous leadership and conflict resolution—sustained alongside stable personality traits. The findings confirm that ChatAgents-driven agents successfully simulate sophisticated group dynamics, characterized by natural turn-taking, contextual relevance, and strategic decision-making.

The principal contributions of this paper are summarized as follows:

- **Self-Driven Order:** We propose a novel multi-agent interaction paradigm that enables agents to autonomously decide when to speak based on contextual and objective-driven reasoning.
- **The ChatAgents Architecture:** This framework integrates a context-driven module activator with a preemptive speaking mechanism to support human-like autonomous conversational decisions.

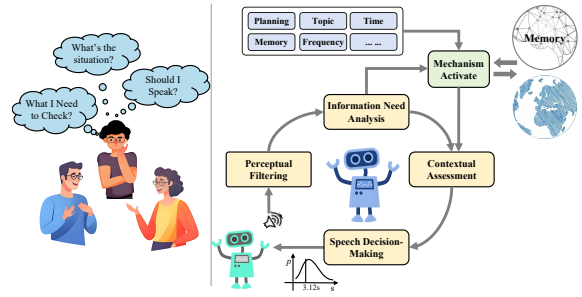


Fig. 2. The proposed framework. The design revolves around three core questions: “What do I need to check?”, “What is the situation?”, and “Should I speak?”. Based on these, the agent perceives the environment, analyses the current state, activates relevant mechanisms to acquire information, evaluates the situation to determine speech readiness, and finally competes for speaking opportunities.

- **A Comprehensive Validation Platform:** We demonstrate the efficacy of our framework through extensive scenario simulations and a multi-dimensional evaluation protocol, thereby advancing research toward more human-like autonomous systems.

II. RELATIVE WORK

Research on LLM agent autonomy has rapidly expanded from single-agent to multi-agent collaboration. Multi-agent systems achieve superior collective intelligence through memory sharing, distributed planning, and collaborative tool use, exhibiting dialectical thinking in debates [13] and high-quality creative output via role specialization [14]. This signifies entry into a “social cognition” phase, where the key challenge is balancing individual autonomy with group synergy [15], requiring both algorithmic innovation and practical methodologies for efficient coordination [16]. The framework proposed in this study extends agent autonomy boundaries. Beyond the standard “Memory-Reflect-Plan-Act” paradigm, it fully automates cognitive processes (memory retrieval, goal induction, topic analysis) and interactive behaviors (e.g., speech timing). This enables agents to compete for speaking opportunities based on real-time context, more closely replicating human social dialogue logic.

Designing effective interaction mechanisms for multi-agent collaboration remains a key challenge. Current approaches primarily follow four paradigms—Rule-Based Order, Centralised Order, Random Order, and Designated Order—each with notable limitations. Rule-Based and Designated Orders enforce predetermined or speaker-chosen sequences, which can restrict adaptability and marginalise less active agents [17]. Centralised Order introduces coordination overhead and suppresses local autonomy [18], while Random Order encourages diversity but often leads to incoherent dialogue [19]. Common across these methods is their limited responsiveness to conversational dynamics such as urgency, importance, or social norms, which constrains their utility in open-domain and socially realistic multi-agent interactions.

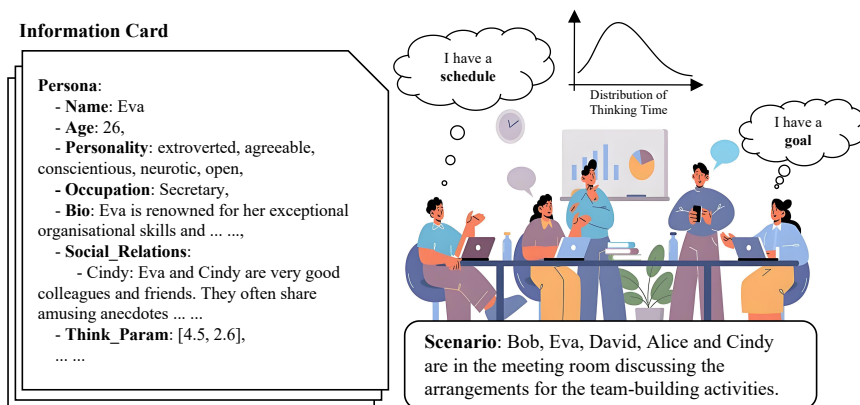


Fig. 3. Group Chat Scenario Settings. By configuring scenario descriptions, character profiles, personalities, objectives, and response time distributions, a group chat simulation can be autonomously conducted in ChatAgents.

III. METHODOLOGY

To address the core challenges in multi-agent conversation systems—simulating human-like cognitive processes and establishing natural conversational dynamics—our framework integrates the structured cognitive architecture with the competitive speaking mechanism, enabling organic and coherent group conversations.

A. Cognitive Architecture

The simulation of multi-agent group conversations employs a structured cognitive process for message handling. Inspired by Generative Agents [20], each agent first applies perceptual filtering to prioritize information congruent with its designated role, traits, and emotional state, thereby simulating human attentional focus. It then analyzes informational gaps by diagnosing the conversational context and dynamically activating relevant cognitive mechanisms—such as memory retrieval or goal-driven reasoning—to acquire necessary knowledge. Subsequently, the agent performs a contextual assessment to synthesize and interpret the accumulated information, forming a coherent understanding of the discourse. This sequence—information collection prior to interpretation—mirrors typical human reasoning. Finally, based on this comprehension, the agent makes a speech decision, determining whether to contribute to the dialogue while weighing its role, objectives, and the evolving conversational dynamics.

B. Speaking Mechanism

To simulate natural turn-taking without centralized coordination, we propose a dynamic competition mechanism integrating willingness screening and temporal modeling for implementation details). In each conversational round, agents first undergo willingness screening, evaluating their self-desire to speak based on contextual factors—such as topic relevance and emotional tone—and individual factors, including role objectives and personality traits. Only agents whose willingness exceeds a predefined threshold proceed, thereby simulating selective engagement in human discussions.

Agents that pass screening then enter a temporal competition phase. Each qualified agent is assigned a response delay sampled from a personality-specific probability distribution, for instance, proactive roles are typically assigned shorter expected delays. The agent with the shortest delay wins the speaking turn, emulating real-time competition in lively human conversations. To prevent permanent exclusion of slower agents and to model cognitive persistence, response times are dynamically adjusted based on conversational continuity.

The mechanism robustly accommodates edge cases: if multiple agents share the shortest delay, one is randomly selected to maintain flow; if no agent passes willingness screening, the system enters a brief waiting period before re-evaluation. By integrating context-aware willingness assessment with moderated temporal competition, our approach enables agents to exhibit human-like, coherent conversational dynamics beyond rigid turn-taking schemes.

Further framework details are available at [ChatAgents](#), containing parameter values and detailed formulations for willingness screening, delay sampling, and dynamic adjustment.

IV. EXPERIMENTS

A. Objectives and Setup

This study aims to validate whether the proposed framework enables multi-agent systems to autonomously generate logically coherent, character-consistent, and narratively closed group conversations based solely on predefined personality settings, scenario descriptions, and an internal memory mechanism, without human intervention. To validate the framework, a team-building scenario was simulated in which five agents collaboratively planned an event within a 30-minute meeting. Each agent was assigned a distinct personality based on a binary Big Five model, which was expanded into ten behavioural descriptions via a large language model to guide conversational choices. The scenario tests the framework’s ability to model time-constrained group dynamics while maintaining behavioural consistency.

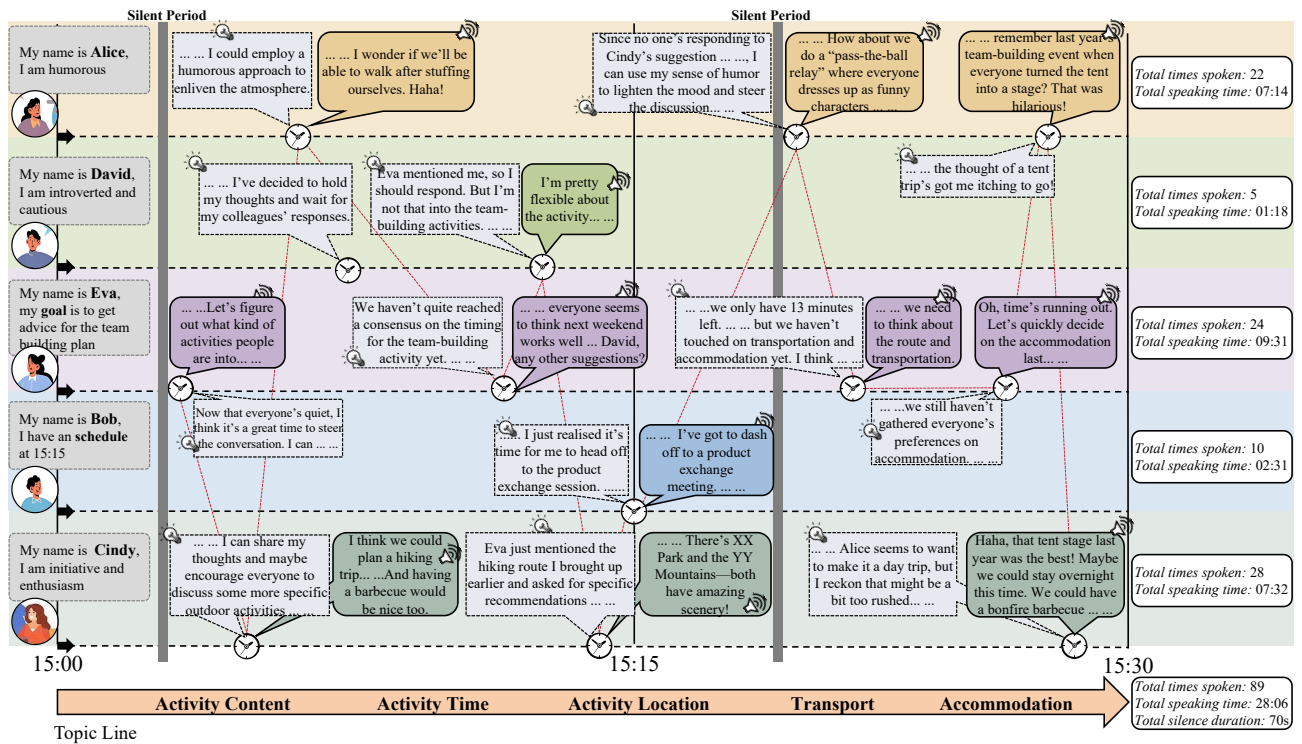


Fig. 4. Example of key conversations in team - building activity scenarios. The conversations and thoughts of the characters at critical time points are presented in chronological order. The participants exhibited diverse interaction styles: Alice was humorous and witty, Bob became increasingly urgent due to time constraints, Cindy proactively engaged without an assigned agenda, David was introverted and cautious, and Eva assumed a facilitative role to secure concrete outcomes for the team-building plan.

B. Results

Overall Dialogue Analysis: In the simulated team-building discussion scenario (Figure 4), each participant exhibited distinct and autonomous interaction patterns that emerged naturally from their predefined personality traits and objectives. These behaviours—ranging from humor and facilitation to urgency and proactive engagement—collectively created a dynamic, multi-turn conversation. The simulation illustrates how agent autonomy fosters diverse roles and yields authentic group dynamics absent external scripting.

Character Personality Consistency Test: The personality traits of simulated characters were evaluated using GPT-4o to complete the Big Five Inventory (BFI) based on their dialogue context, as in Table I. The results confirm that characters generated by the ChatAgents framework stably exhibited statistically significant differences ($P < .001$) in all five personality dimensions, aligning with their predefined settings. This demonstrates the framework’s capability to autonomously maintain distinct and consistent personality profiles during multi-agent conversations.

Statistical analysis: Table II quantifies the behavioral patterns and mechanism selection strategies of five autonomous agents during a 30-minute simulation. The results demonstrate that each agent autonomously employs specific socio-cognitive mechanisms aligned with its predefined personality and objectives. Governed by distinct personality traits and situ-

TABLE I
EVALUATION RESULTS OF THE BIG FIVE PERSONALITY TRAITS.

Trait Dimension	Trait Description	Mean [95% CI]	P
Extraversion	Extroverted	4.80 [4.61, 4.99]	<.001
	Introverted	3.75 [3.30, 4.20]	
Agreeableness	Agreeable	4.85 [4.62, 5.08]	<.001
	Antagonistic	3.90 [3.70, 4.10]	
Conscientiousness	Conscientious	4.67 [4.50, 4.84]	<.001
	Unconscientious	3.03 [2.91, 3.15]	
Neuroticism	Neurotic	2.15 [1.88, 2.42]	<.001
	Emotionally Stable	1.33 [1.15, 1.51]	
Openness	Open	4.60 [4.19, 5.01]	<.001
	Closed	3.92 [3.70, 4.15]	

Note. This table presents the scoring profiles of individuals across five major personality dimensions. Under each dimension, two tendencies are described. For positive tendencies, a score closer to 5 is better, while for negative tendencies, a score closer to 1 is better. The Neuroticism dimension is the exception.

ational contexts, agents manifested varied interaction dynamics—ranging from utterance frequency and silence episodes to the strategic activation of memory retrieval and topic analysis. These diverse behaviors collectively orchestrated the group discussion flow, validating the capacity of ChatAgents to facilitate nuanced, personality-driven social interactions in collaborative environments.

TABLE II
STATISTICAL ANALYSIS OF AGENT BEHAVIOURS AND MECHANISM SELECTION IN SIMULATED GROUP CHAT INTERACTIONS.

Character	Personality Traits	Special Settings	Spoken	Wanting to Speak	Silent	Mechanism (On-Off)					Wanting to Speak After Silence	Speaking After Silence
						Memory Retrieval	Goal Summary	Topic Analysis	Time Analysis	Speech Count Stats.		
Alice [3.6, 3.0]	introverted, agreeable, unconscious, neurotic, open	tend to value Bob's opinions more	8	11	108	82-37	21-98	0-119	56-63	71-48	2	1
Bob [4.3, 2.5]	extroverted, agreeable, unconscious, emotionally stable, closed	have another engagement at 3:15 p.m.	18	28	91	12-107	35-84	8-111	119-0	35-84	9	5
Cindy [3.9, 2.8]	extroverted, antagonistic, unconscious, emotionally stable, open	not comfortable with silence in a conversation.	29	36	83	27-92	17-102	0-119	64-55	110-9	15	13
David [4.8, 2.0]	introverted, antagonistic, unconscious, emotionally stable, closed	not particularly interested in this team-building activity	10	18	101	57-62	4-115	5-114	61-58	53-66	1	0
Eva [4.5, 2.6]	extroverted, agreeable, conscientious, neurotic, open	hope to obtain information regarding the time, location, format, and content of the team-building activity	15	20	99	46-73	60-59	34-85	90-29	94-25	3	3

Special Behavior: In the team-building discussion, agents exhibited key emergent social behaviors that demonstrate advanced autonomy. Eva acted as an emergent leader by steering the conversation from details to core goals. A conflict between Bob (safety-focused) and Cindy (nature-focused) was constructively mediated when Cindy proposed a compromise solution. Additionally, Cindy generated innovative proposals by synthesizing others' ideas, showcasing creative autonomy. These behaviors illustrate how agents can dynamically manage discussions, resolve disagreements, and foster innovation.

For additional experimental details and a comprehensive description of the methodology, please refer to [ChatAgents](#).

V. CONCLUSIONS AND LIMITATIONS

We presented ChatAgents, a framework leveraging Self-Driven Order to facilitate context-aware autonomy in multi-agent systems. Experimental results validate that agents can effectively replicate complex human social dynamics, such as spontaneous leadership and conflict resolution, without compromising character consistency. While LLM-inherent randomness and limited interaction mechanisms remain constraints, this work provides a foundation for more sophisticated social modeling. Future efforts should aim to diversify adaptive strategies and mitigate hallucinations to improve framework reliability in large-scale scenarios.

REFERENCES

- [1] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, *et al.*, "Improving language understanding by generative pre-training," *Technical report, OpenAI* **0**, 2018.
- [2] S. Wang, M. Hu, Q. Li, M. Safari, R. L. Qiu, and X. Yang, "Capabilities of gpt-5 on multimodal medical reasoning," in *Medical Imaging 2026: Imaging Informatics*, **13930**, pp. 77–86, SPIE, 2026.
- [3] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, *et al.*, "Llama 2: Open foundation and fine-tuned chat models," *arXiv preprint arXiv:2307.09288* **0**, 2023.
- [4] Z. Yang, E. Li, J. He, A. Priyanshu, B. Saglam, P. Kassianik, S. Weerawardhena, A. Vellore, B. Nelson, N. Javidnia, *et al.*, "Llama-3.1-foundationai-securityllm-reasoning-8b technical report," *arXiv preprint arXiv:2601.21051*, 2026.
- [5] F. Jiang, C. Pan, L. Dong, K. Wang, M. Debbah, D. Niyato, and Z. Han, "A comprehensive survey of large ai models for future communications: Foundations, applications and challenges," *IEEE Communications Surveys & Tutorials*, 2026.
- [6] J. Hu, W. Zhang, Y. Wang, Y. Hu, B. Xiao, M. Tan, and Q. Du, "Dynamic compressing prompts for efficient inference of large language models," *arXiv preprint arXiv:2504.11004*, 2025.
- [7] R. Buyya *et al.*, "Agentic artificial intelligence (ai): Architectures, taxonomies, and evaluation of large language model agents," *arXiv preprint arXiv:2601.12560*, 2026.
- [8] C. Du, X. Wang, A. Chen, W. Li, R. Xu, J. Liu, Z. Huang, R. Tian, Z. Sun, Y. Li, *et al.*, "Her: Human-like reasoning and reinforcement learning for llm role-playing," *arXiv preprint arXiv:2601.21459*, 2026.
- [9] J. Zheng, C. Li, P. Wu, and L. Pan, "Rolesimllm: Towards large-scale and comprehensive social propagation simulation via role-based llm-driven agents," *Information Processing & Management* **63**(5), p. 104689, 2026.
- [10] W. Fang, L. Li, K. Shi, Y. Yang, and J. Zhang, "Belief-driven multi-agent collaboration via approximate perfect bayesian equilibrium for social simulation," *arXiv preprint arXiv:2603.24973*, 2026.
- [11] L. Liu, L. Wang, W. Niu, and S. Hua, "Dynamic sanctioning mechanism for cooperative multi-agent systems," *Expert Systems with Applications* **296**, p. 128873, 2026.
- [12] Z. Meng, Z. Xie, Z. Ye, C. Lu, Z. Liu, Z. Niu, Y. Hu, and S. Cao, "Gca-agent: Enhancing group chat communication through dialogue agents system," *arXiv preprint arXiv:2603.05240*, 2026.
- [13] M. Zhang, J. Kim, S. Xiang, J. Gao, and C. Cao, "Dynamic role assignment for multi-agent debate," *arXiv preprint arXiv:2601.17152*, 2026.
- [14] H. Lim, D. Choi, S. Nam, B. Kim, and H. Hong, "Understanding human-multi-agent team formation for creative work," *arXiv preprint arXiv:2601.13865*, 2026.
- [15] Y. Zhao, W. Zhang, A. Rechkemmer, A. Meroño-Peñuela, and E. Simperl, "Exploring the design of genai-based systems to support socially shared metacognition," *arXiv preprint arXiv:2603.08894*, 2026.
- [16] B. Turkkan, E. Rodrigues, T. Kosar, A. Charapko, A. Ailijiang, and M. Demirbas, "How to evaluate distributed coordination systems?—a survey and analysis," *IEEE Transactions on Parallel and Distributed Systems* **37**(1), pp. 198–212, 2025.
- [17] G. Li, H. Hammoud, H. Itani, D. Khizbullin, and B. Ghanem, "Camel: Communicative agents for "mind" exploration of large language model society," *Advances in Neural Information Processing Systems* **36**, pp. 51991–52008, 2023.
- [18] C. Qian, W. Liu, H. Liu, N. Chen, Y. Dang, J. Li, C. Yang, W. Chen,

Y. Su, X. Cong, *et al.*, “Chatdev: Communicative agents for software development,” in *Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers)*, pp. 15174–15186, 2024.

- [19] B. S. Manning, K. Zhu, and J. J. Horton, “Automated social science: Language models as scientist and subjects,” tech. rep., National Bureau of Economic Research, 2024.
- [20] J. S. Park, J. O’Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, “Generative agents: Interactive simulacra of human behavior,” in *Proceedings of the 36th annual acm symposium on user interface software and technology*, pp. 1–22, 2023.