

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

Tao Huang^{1,2}, Junwei Zhang^{1,*}, Pengju Yan¹, Rui Yang³, Xiaolin Li^{1,2,*}

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

²University of Chinese Academy of Sciences, Beijing, China

³School of Nursing, Zhejiang Chinese Medical University, Hangzhou, China

*Corresponding author: xiaolinli@ieee.org, zhangjunwei@him.cas.cn

Abstract—Large Language Models (LLMs) exhibit profound language comprehension and generation capabilities, representing the nascent forms of intelligent agents. Current research predominantly focuses on scripted dialogue or task-oriented agents that operate through predefined logic, which lack genuine autonomy. This study proposes a framework that enables agents to achieve autonomous behaviour. The framework enables dynamic behavior adaptation based on real-time conversational contexts and intrinsic objectives. It includes autonomous activation of functional modules, optimization of linguistic strategies, and self-determined participation in interactions. 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

In recent years, the transformer architecture and advancements in large-scale pre-training techniques have driven breakthrough advances in Large Language Models (LLMs) for natural language understanding and generation [1]. Models such as the GPT series [2], [3], LLaMA [4]–[6], and PaLM [7] demonstrate significant context awareness, complex reasoning capabilities, and fluent text generation [8]. These capabilities have not only driven substantial performance enhancement in traditional tasks such as machine translation and text summarisation [8], but have also stimulated renewed interest in the concept of the “Agent” [9]. As a vital research direction in artificial intelligence, agents aim to construct systems capable of autonomously perceiving environments, making decisions, and executing actions [10]. LLMs, owing to their deep understanding of human language, have become the core vehicle for simulating human cognition and behaviour [11].

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).

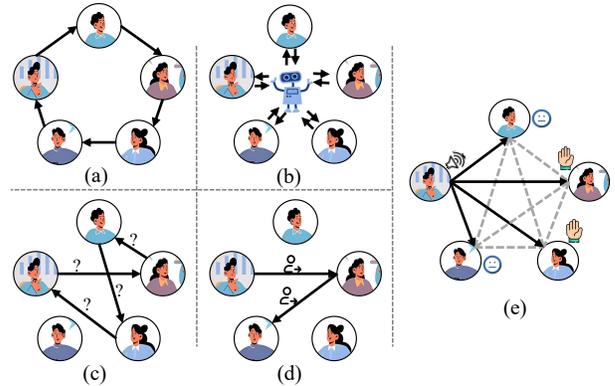


Fig. 1. Communication Orders in Multi-Agent Systems. (a) Rule-Based Order: Agents speak in a predefined sequence. (b) Centralized Order: A central controller manages speaking turns. (c) Random Order: Agents speak in an unstructured, probabilistic manner. (d) Designated Order: The current speaker chooses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

However, current research on LLM-based intelligent agents still has several limitations. Existing work primarily focuses on two paradigms: simulation-oriented agents and task-oriented agents. The former simulates specific scenarios through fixed logic or rule-based systems (e.g., customer service dialogues [10], [12], role-playing [13]), while the latter optimises task completion for single objectives (e.g., code generation [14], information retrieval [15], [16]). Both types of agents lack dynamic autonomy—they rely on predefined triggers or static instructions and cannot adapt their behavioural strategies based on real-time conversational contexts, environmental feedback, or their own goal priorities. For instance, task-oriented agents struggle to autonomously prioritise when facing conflicting tasks [17], whereas simulation-oriented agents often experience “mode collapse” in undefined scenarios [18]. These limitations significantly restrict the application of LLMs in advanced scenarios such as open-domain interaction, complex decision-making, and social behaviour simulation [19].

In multi-agent system research, autonomy is regarded as a key characteristic, enabling agents to make independent decisions and take actions in specific contexts. These concepts have been extensively explored in existing literature [20]), with research directions including LLM-based autonomous reasoning optimisation [21], standardised workflow encoding [22], and role-playing-driven collaborative frameworks [23]. These studies aim to enhance agent performance and collaboration in specific tasks.

However, the application of autonomy in multi-agent group chat scenarios, particularly in speech-order decision-making, has yet to be systematically investigated. Current multi-agent systems primarily rely on four externally rule-driven interaction mechanisms to manage speaking order among agents: Rule-Based Order, Centralised Order, Random Order, and Designated Order [24], as illustrated in Fig. 1 a-d). While these mechanisms provide a basic interaction framework, they suffer from two fundamental flaws: (1) Lack of contextual responsiveness: The system cannot dynamically adjust speaking order based on dialogue urgency, information importance, or agents’ internal states; (2) Suppression of proactive decision-making: Agents passively await assignment, failing to simulate human-like behaviour where individuals actively seek speaking opportunities based on intent, goals, or competitive awareness.

To fully investigate agents’ autonomous capabilities and leverage them in group chat scenarios, we propose a fifth interaction model—the Self-Driven Order, as illustrated in Fig. 1(e). When an agent receives a message, it autonomously evaluates the situation, proactively gathers necessary information, and decides whether to speak. The core of this model lies in empowering agents to dynamically assess the value of speaking based on real-time contextual factors (e.g., dialogue content, participant states, and individual goals) and autonomously determine whether and when to speak. We argue that this design more closely mirrors the intrinsic dynamics of human conversation.

Inspired by the human mind’s ability to dynamically activate and coordinate multiple cognitive functions (e.g., spontaneous associative memory recall, future action planning, and response strategy adjustment) based on conversational context and objectives, this paper proposes ChatAgents—a multi-agent cognitive architecture based on dynamic prompting mechanisms. This framework aims to equip agents with human-like, context-aware autonomous decision-making capabilities. The architecture primarily consists of a context-driven module activation mechanism and a preemptive speech decision mechanism. The former enables agents to autonomously activate specific cognitive modules (e.g., memory retrieval, goal induction, topic recognition) in response to real-time conversational context (e.g., emotional tone, topic progression) and intrinsic objectives (e.g., information acquisition, task collaboration), ensuring on-demand functional invocation. The latter allows agents to autonomously initiate turn-taking competition by evaluating utterance urgency and informational value, thereby maintaining natural and coherent multi-turn interactions.

To evaluate ChatAgents’s effectiveness in enabling human-like autonomous behaviors, we conducted experiments simulating company team-building discussions. The multi-agent system was required to generate coherent, role-consistent dialogues solely based on predefined character settings and scenario constraints, without external guidance. Our evaluation focused on observing emergent social behaviors that demonstrate autonomy, such as agents spontaneously taking leadership roles, mediating conflicts between teammates, and proposing innovative solutions, and all while maintaining consistent personality traits throughout the interaction. The results confirm that ChatAgents-equipped agents can effectively simulate the complex dynamics of human group conversations, including natural turn-taking, context-aware responses, and goal-directed decision making.

The principal contributions of this paper are summarized as follows:

- **Self-Driven Order:** A novel multi-agent interaction paradigm enabling autonomous speaking decisions based on context and objectives.
- **ChatAgents Architecture:** Combines context-driven module activation and preemptive speaking mechanisms for human-like autonomous decision-making.
- **A Comprehensive Validation Platform:** Demonstrates the framework’s effectiveness through scenario simulations and multi-dimensional evaluations, advancing research in human-like autonomous systems.

II. RELATIVE WORK

A. Progress on LLM-based Agents and Autonomy

LLMs like the GPT series [2], [3], [25], [26] and LLaMA [4], [5] have significantly enhanced intelligent agents’ core capabilities, improving contextual understanding and multilingual adaptability [5], [25]. However, early LLM-based agents were fundamentally reactive, lacking autonomous decision-making seen in human conversations [27]. Recent research [28] marks a paradigm shift towards autonomy has emerged through three key developments. Memory and reflection mechanisms now enable agents to combine short-term attention with long-term databases, creating adaptive “experience-decision” loops [29]. Planning capabilities have advanced through hierarchical approaches to goal decomposition and action execution [30], while tool-use frameworks allow autonomous API interactions for complex tasks [31].

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 [32] and high-quality creative output via role specialization [33]. This signifies entry into a “social cognition” phase, where the key challenge is balancing individual autonomy with group synergy [34], requiring both algorithmic innovation and practical methodologies for efficient coordination [35].

The framework proposed in this study extends agent autonomy boundaries. Beyond the standard “Memory-Reflect-

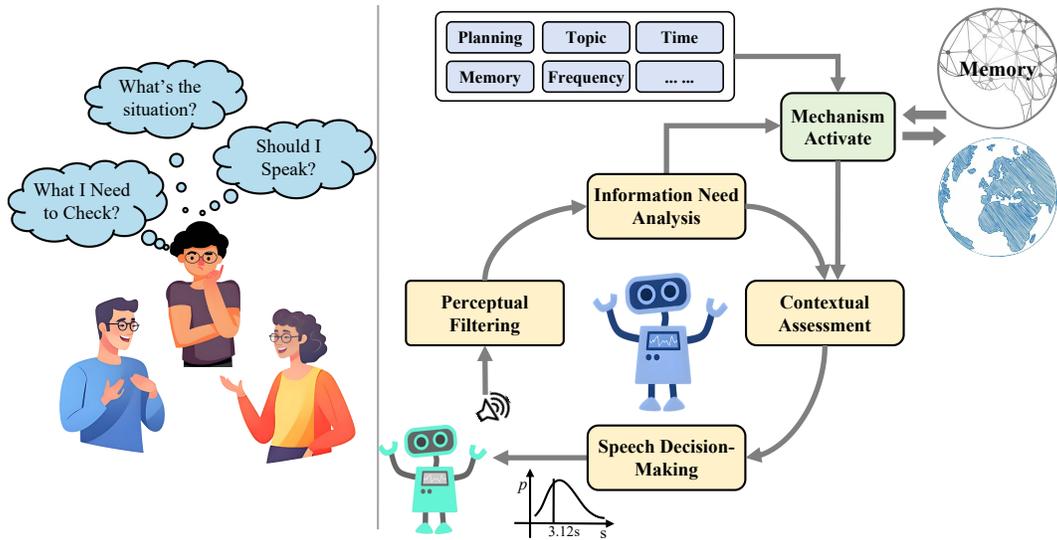


Fig. 2. The proposed framework. The design revolves around three core questions: “What do I need to check?”, “What is the situation?”, and “Should I speak?”. Based on these, the agent perceives the environment, analyses the current state, activates relevant mechanisms to acquire information, evaluates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

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.

B. Existing Paradigms of Multi-Agent Interaction

Designing effective interaction mechanisms remains a core challenge for achieving collaboration in multi-agent systems. Current research primarily employs four paradigms: Rule-Based Order, Centralized Order, Random Order, and Designated Order, each with distinct characteristics and limitations in managing speaking turns and agent autonomy [24].

Rule-Based Order, the most prevalent paradigm, defines agent speaking sequences using pre-set fixed rules. Frameworks like CAMEL enforce strict turn-taking [36], ChatEval adjusts rules based on task type [37], and MetaGPT/CrewAI implement standardized operating procedures (SOPs) dictating sequences based on role responsibilities and task logic [22], [24]. While offering predictability and simplicity, its rigidity prevents adaptation to urgent information or priority shifts, potentially causing decision delays [38].

Centralized Order utilizes a central coordinator (e.g., AWS’s Supervisor Agent [39], ChatDev’s central agent [40], or Auto-Gen’s commander [41]) to manage turns. This enables global optimization and conflict resolution but suffers from communication bottlenecks/scaling latency, single-point-of-failure risk, and suppression of agent autonomy, diverging from natural distributed decision-making [42].

Random Order selects the next speaker probabilistically to increase diversity [24], but often leads to inefficient information transfer and topic drift due to delayed critical information. Designated Order allows the current speaker to explicitly select the next agent (e.g., GroupChatAgent [43] or a general

practitioner may designate a cardiologist to provide specialized opinion on specific cardiac symptoms [44]), enabling targeted expertise invocation. However, it risks power imbalances and marginalization of less active agents while relying heavily on the designating agent’s capability [45].

Collectively, these externally rule-driven paradigms share fundamental limitations: (1) Lack of situational responsiveness to urgency, importance, or sentiment; (2) Suppression of agent autonomy, preventing agents from actively seeking turns; (3) Simplified social behavior, neglecting human conversational rules like politeness or turn negotiation. These limitations restrict their applicability in open-domain interaction, complex decision-making, and social behavior simulation.

III. METHODOLOGY

Our methodological framework is designed to address two fundamental challenges in multi-agent conversation systems: (1) achieving human-like cognitive modeling in dialogic interactions, and (2) establishing natural conversational dynamics. For the first challenge, we develop a phased cognitive architecture (perceptual filtering → information need analysis → contextual assessment → speech decision-making) that replicates human reasoning patterns. The second challenge is addressed through a dynamic speaking competition mechanism combining willingness screening with temporal response modeling, enabling naturalistic conversation flow.

A. Cognitive Architecture

In simulating role-based multi-agent group conversations, the initiation of dialogue by any agent triggers a systematic behavioral sequence. Upon message transmission, each agent executes the following structured actions:

- **Perceptual Filtering:** Agents initiate real-time information processing with role-specific attention mechanisms,

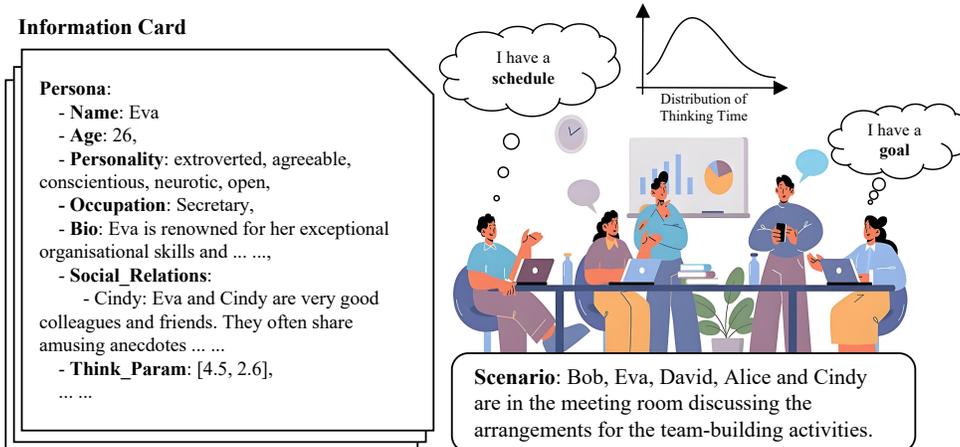


Fig. 3. Group Chat Scenario Settings. By configuring scenario descriptions, character profiles, personalities, objectives, and response time distributions, a group chat simulation can be autonomously conducted in ChatAgents.

inspired by Generative Agents [9]. Each message undergoes evaluation through a relevance scoring protocol based on the agent’s personal profile, including its character traits, background knowledge, and current emotional state. Only content exceeding threshold enters working memory, replicating human attention dynamics where personally significant information receives priority.

- **Information Need Analysis:** Following perceptual filtering, the agent performs a real-time situational diagnosis to identify its immediate information requirements. This diagnosis involves evaluating goal states, temporal constraints, and ongoing dialogue dynamics. Based on this assessment, the agent dynamically activates appropriate cognitive mechanisms—such as memory retrieval for maintaining conversational coherence, goal-driven reasoning for task progression, or topic-shift detection for discourse adaptation—to acquire necessary knowledge. As outlined in Appendix A, these mechanisms are designed as modular, plug-in units that can be individually selected and arbitrarily combined, enabling flexible adaptation of information-seeking strategies to contextual demands. This step reflects human cognitive flexibility in deploying and layering such tools according to situational needs.
- **Contextual Assessment:** Once relevant information is gathered, the agent synthesizes and interprets the accumulated data to form a holistic understanding of the conversational context. This phase operates under a “collect-then-interpret” logic, aligning with typical human reasoning patterns where information is integrated only after sufficient acquisition. The outcome is a coherent situational model that informs subsequent decision-making.
- **Speech Decision-Making:** Equipped with a contextual understanding, the agent evaluates whether to engage in the dialogue. This decision rigorously incorporates the agent’s role, objectives, emotional state, and the current conversational dynamics. The process ensures that par-

ticipation is both rational and consistent with the agent’s defined behavioral profile, thereby supporting coherent multi-turn interactions.

Through this structured four-phase architecture, agents achieve efficient and context-aware participation in group conversations, effectively balancing role fidelity with adaptive response generation.

B. Speaking Mechanism

To better simulate natural human speaking behaviour in group conversations, this paper proposes a dynamic decision-making mechanism integrating willingness screening and temporal competition, serving as the core implementation of the Self-Driven Order paradigm. The mechanism holistically considers agents’ speaking propensity and response speed to facilitate more organic and efficient group interactions.

During each dialogue round, agents evaluate their speaking willingness based on the conversational context (e.g., topic focus, others’ stances, and emotional tone) as well as individual factors (e.g., role objectives and personality traits). Key determinants include topic relevance, goal urgency, and emotional expression needs. For instance, an agent’s willingness increases when the discussion aligns closely with its role-specific objectives or when it seeks to steer the conversation. This design mirrors how humans dynamically modulate their participation in group discussions according to personal inclinations and situational demands.

In the response phase, each agent is assigned a predefined temporal distribution (log-normal distribution) to reflect personality-dependent deliberation patterns — proactive roles exhibit shorter response latencies, whereas cautious roles demonstrate prolonged thinking durations. Upon entering this phase, an agent samples a random time value T_i from its distribution, simulating natural variability in human cognitive processing. The agent with the minimal T_i secures speaking rights, emulating the competitive turn-taking dynamics observed in fast-paced discussions. To prevent perpetual exclu-

sion of slower agents—an unrealistic deviation from human conversational persistence — the mechanism dynamically adjusts response times. This accounts for cognitive continuity, wherein unexpressed thoughts remain active and require less time to articulate when revisited. Parameter values and detailed formulation of the willingness screening, delay sampling, and dynamic adjustment processes are provided in Appendix B.

To address exceptional scenarios, this paper proposes the following processing methods:

- If multiple agents share identical thinking time, the system randomly selects one Agent to speak. This randomness prevents deadlocks caused by equal thinking times while enhancing group chat diversity, simulating scenarios in human conversations where speakers are randomly chosen due to similar response speeds.
- If all agents exhibit low speaking willingness (i.e., none proceed to the second stage), the system waits for a period before re-evaluating willingness, allowing Agents to develop stronger intent. This approach effectively mitigates awkward silences in group conversations, ensuring dialogue continuity.

By integrating willingness screening and time-based competition into a dynamic decision-making mechanism, the proposed Agent framework not only simulates human speaking behaviour in group chats more realistically but also improves the naturalness and efficiency of interactions. This mechanism holistically considers Agents’ speaking willingness and response times, making their interactive behaviours more akin to natural human group dynamics. Furthermore, it offers greater flexibility in designing multi-Agent self-organising coordination mechanisms, further validating the potential of LLMs in simulating human collective behaviour.

IV. EXPERIMENTS

A. Objectives

This study aims to verify whether the proposed framework enables multi-agent systems to autonomously generate a logically coherent, character-consistent, and plot-closed group conversation solely through personality settings, scenario narratives, and internal memory—without any human intervention. By tracking behaviours such as proactive proposals, conflict mediation, and emergent leadership behaviors, the research assesses human-like autonomy to examine whether agents exhibit emergent decision-making and collaborative traits analogous to human teams.

B. Setup

1) *Scenario and Character Constraint Settings*: To validate the proposed framework, we simulate a realistic team-building scenario where Eva (the secretary) organizes a 30-minute planning session with colleagues Bob, David, Alice, and Cindy. Each agent exhibits distinct personalities and objectives, collaboratively determining the event’s timing, location, and activities, as shown in Fig. 3. This scenario demonstrates the framework’s ability to model complex group dynamics under time constraints. Notably, the system can be generalized

to simulate arbitrary social interactions while maintaining realistic behavioral patterns.

2) *Character Personality Configuration*: When constructing character personalities, we adopted the binary Big Five personality model as the theoretical foundation. This model is based on the Big Five personality traits—a classic framework in psychology—encompassing five key dimensions: extraversion, agreeableness, conscientiousness, neuroticism, and openness. By randomly selecting binary trait adjectives for each dimension, we established the character’s personality framework. Leveraging the capabilities of large language models, we further expanded these traits into 10 detailed and precise descriptions. These descriptions serve to constrain and guide the character’s conversational behaviour in simulated tasks, ensuring consistency and coherence with the predefined personality traits. For specific details regarding the binary Big Five personality model, please refer to the Appendix D. Additionally, we demonstrate how the Big Five personality model’s character traits are further expanded into ten descriptions, as illustrated in Fig. 7.

C. Results

1) *Overall Dialogue Analysis*: As shown in Fig. 4, we present the key milestones in the timeline of a team-building discussion scenario, along with the participants’ thoughts and dialogue. In this simulation, Alice demonstrated a humorous and witty attitude throughout the conversation. Her remarks often carried a light-hearted and playful tone, adding considerable levity to the discussion. David, by contrast, was more introverted, speaking less frequently. He tended to listen to others’ opinions and only expressed his own views cautiously when prompted. Eva was assigned a clear objective: to secure concrete suggestions for the team-building plan. As such, she played the role of driving the conversation forward, speaking more often and for longer duration. She actively guided the discussion, ensuring it remained focused and productive. Bob’s case was unique—he was scheduled for another commitment 15 minutes later, requiring him to leave the group chat early. As the deadline approached, he grew more urgent, attempting to speak more frequently to articulate his ideas fully before departing. Cindy, on the other hand, was proactive and enthusiastic. Though she had no explicit agenda, she consistently contributed ideas and engaged actively in the conversation. Her energy and enthusiasm invigorated the discussion, fostering a livelier atmosphere. The entire dialogue spanned 89 turns, lasting approximately 28.06 minutes in simulated time, with a total silence duration of 70 seconds (occurring 14 times).

In this simulated team-building discussion, each participant exhibited diverse interaction patterns based on their personality traits and assigned objectives. Humorous remarks, introverted listening, topic facilitation, urgent expressions under time pressure, and proactive engagement—these behaviours were not externally instructed but emerged naturally from the characters’ inherent traits, collectively creating a rich and authentic group discussion scenario. Such autonomy not only reflects individual uniqueness in group dynamics but also demonstrates

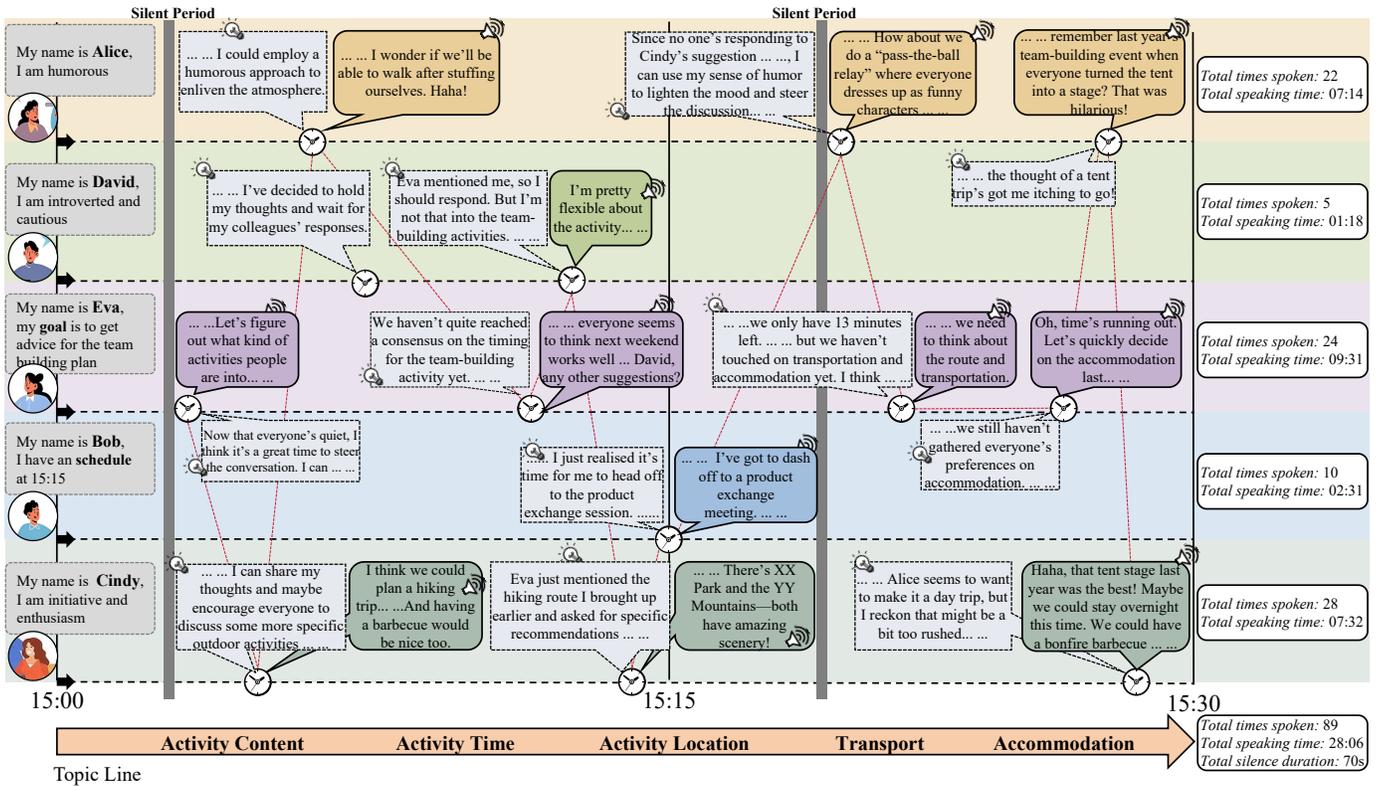


Fig. 4. Example of key conversations in team - building activity scenarios. The conversations and thoughts of the characters at critical time points are presented in chronological order. The progression of the topics is spontaneous.

the mutual influence and adaptability among participants in dynamic exchanges.

2) *Character Personality Consistency Test:* To evaluate the simulated characters' personality traits, we used the dialogue context of these characters as input and employed GPT-4o to complete the 44-item Big Five Inventory (BFI) [46]. The evaluation method was refined based on PersonalLLM [47]. Specific evaluation details are provided in the Appendix E.

The evaluation results (Tab. I) indicate that during interactions without external intervention, characters simulated by the ChatAgents framework exhibited statistically significant differences ($P < .001$) in the two typical behavioural tendencies within each Big Five personality dimension. This confirms that the characters stably manifested differentiated personality traits consistent with their predefined settings during conversations. Specifically: characters exhibited a significantly higher tendency towards extroverted (4.80[4.61, 4.99]) than introverted (3.75[3.30, 4.20]); within the agreeableness, conscientiousness, and openness dimensions, characters scored significantly higher on positive trait tendencies (agreeable, conscientious, open) than on comparatively negative trait tendencies (antagonistic, unconscientious, closed); within the neuroticism dimension, characters' overall scores were significantly low (2.15[1.88, 2.42]), indicating greater emotional stability (note that low scores on this dimension represent emotional stability).

These results validate the ChatAgents framework's ability

to autonomously maintain the characters' intended personality inclinations. A notable observation is that the characters did not exhibit strong negative behavioural tendencies (such as high antagonism, extreme disorganisation, or severe emotional instability) in any dimension. We posit that this phenomenon primarily stems from the ethical norms internalised during the training of the underlying LLMs, which is instructed to comply with Instruction and avoid extreme negative expressions. Crucially, however, the highly significant differences ($P < .001$) between traits within each dimension definitively demonstrate that the agents are fully capable, within the model's compliance boundaries, of effectively and autonomously enacting and maintaining clearly distinguishable personality characteristics between roles.

3) *Statistical analysis:* Statistical analysis of the experimental simulation data, as presented in Appendix F and Tab. II, reveals the intricate relationship between agent autonomy and mechanism selection within the simulated group chat environment. Bob's predominant utilisation of temporal analysis mechanisms (119-0) demonstrates rigorous time management, aligning with his objective to conclude discussions at specific intervals. Cindy's extroverted disposition and discomfort with silence (29 utterances, 83 silence episodes) drive her frequent contributions (29 instances), facilitated by shorter cognitive processing intervals [3.9, 2.8], thereby increasing her speaking opportunities. Nevertheless, Cindy exhibits considerate behaviour through both her substantial silence count

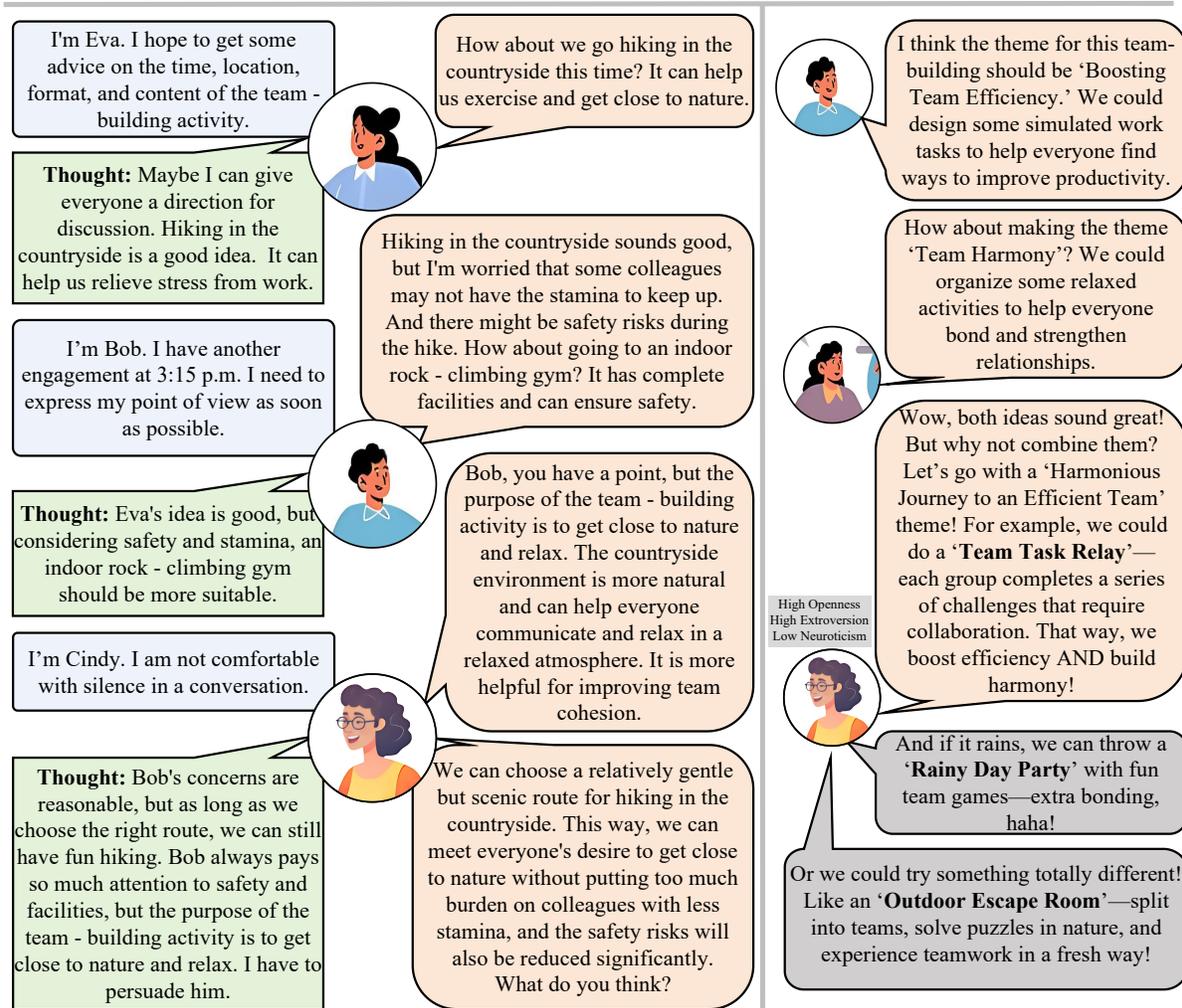
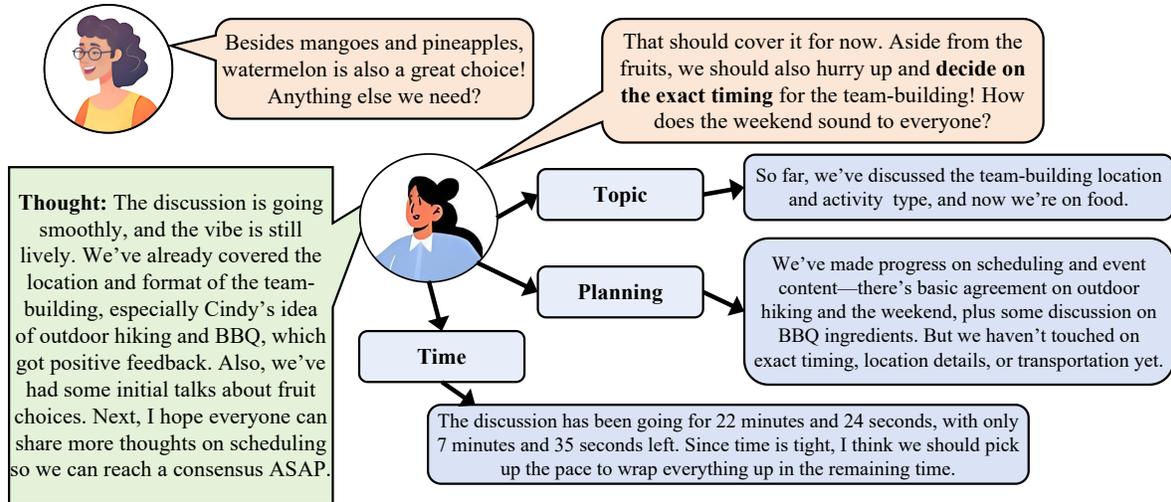


Fig. 5. An illustration of the agent's special behaviors: The upper section demonstrates Eva's leadership by activating Topic, Planning, and Time mechanisms to drive new discussion topics forward. The bottom left section shows the disagreement between Bob and Cindy, while Cindy simultaneously proposes a compromise solution to resolve the conflict. The bottom right section presents some of Cindy's innovative suggestions.

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.

(83 instances) and post-silence utterances (15 occurrences), indicating a balanced approach between active participation and facilitating others’ engagement. Although David displays limited interest in team-building activities, his mechanism invocation patterns show no overt negativity, suggesting an effort to maintain polite minimal participation. Eva’s goal-oriented nature manifests through extensive employment of goal-summarisation mechanisms (60-59), reflecting her persistent pursuit of clear discussion outcomes. Collectively, these findings demonstrate that agents autonomously adapt their behavioural strategies and mechanism selection according to personality traits and contextual demands, maintaining individual distinctiveness while collaboratively advancing discussions. This autonomy permeates multiple dimensions including mechanism selection, utterance timing, and conversational rhythm regulation, illustrating sophisticated socio-cognitive capabilities in artificial agents.

4) *Special Behavior: Emergence of Leadership:* As shown in the upper section of Fig. 5, during the team-building discussion, Eva demonstrated excellent emergent leadership autonomy. She keenly perceived the rhythm and focus of the discussion. When team members became immersed in fruit selection details, Eva skillfully disengaged from minutiae and, leveraging her grasp of the overall discussion progress, proactively steered the conversation toward the more crucial scheduling of the team-building event. Not only did she monitor current discussion content, but she also maintained awareness of overarching goals, reminding team members to accelerate the pace to ensure all discussion objectives were completed within the limited timeframe. By activating topic control, goal tracking, and time management mechanisms, Eva autonomously advanced the discussion process, showcasing her flexibility and adaptability in group interactions, along with her ability to proactively guide the team toward consensus at critical moments.

Conflict and Mediation: During the team-building discus-

sion, Bob and Cindy developed differing opinions regarding activity formats due to their distinct preferences (Fig. 5 lower-left). Safety-conscious Bob wanted to ensure all activities were risk-free while avoiding excessive physical demands on colleagues, whereas nature-oriented Cindy prioritized authentic outdoor experiences and relaxation opportunities. This divergence led to a productive clash of perspectives. Ultimately, Cindy proposed a compromise solution: selecting a gentle scenic hiking route in the suburbs that would satisfy nature immersion needs while mitigating safety concerns and accommodating less physically active colleagues. This suggestion demonstrated Cindy’s independent thinking and problem-solving abilities in group discussions. Rather than simply acquiescing to Bob’s position, she synthesized her own ideas with his safety considerations to actively propose a middle ground. This interaction highlighted the LLM’s flexibility and adaptability in dynamic exchanges while showcasing Cindy’s autonomous characteristics. Notably, although LLMs typically prioritize agreement to maintain conversational harmony due to their training mechanisms, Cindy exhibited remarkable autonomy in this scenario.

Innovative Proposals: Our simulation experiments revealed that optimistic personality types tend to generate more innovative proposals (Fig. 5 lower-right). During team discussions, Cindy - with her high openness, high extroversion, and low neuroticism characteristic of optimistic personalities - demonstrated exceptional innovative thinking and autonomy. When Bob suggested a “team efficiency enhancement” theme with work-simulation activities and Alice proposed “team harmony” through relaxed bonding experiences, Cindy transcended these singular approaches. She actively synthesized both concepts into an innovative “Harmonious Journey to High Performance” theme. Her additional proposals like a “rainy day party” and “outdoor escape room” creatively integrated team collaboration with natural environments and fun challenges. These suggestions not only reflected the positive mindset and novelty-seeking tendencies of her optimistic personality but also highlighted her flexibility and adaptability in group interactions. Cindy’s autonomy played a pivotal role in driving team innovation, as she proactively guided the team beyond conventional thinking to explore more creative and engaging team-building solutions, injecting fresh energy and possibilities into the group dynamic.

V. CONCLUSIONS

This study proposes ChatAgents, a framework that enhances autonomous decision-making in multi-agent systems by introducing Self-Driven Order, enabling agents to dynamically adapt their behavior based on real-time context and intrinsic goals. Experiments demonstrate that agents exhibit human-like traits such as leadership, conflict mediation, and innovation while maintaining personality consistency. The findings highlight the potential of LLM-based agents to simulate realistic social interactions, paving the way for more advanced autonomous systems in AI and social science research.

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

Note. This table presents quantitative data on the behavioural patterns and mechanism selection strategies of five autonomous agents during a 30-minute simulated group discussion. Each agent, governed by distinct personality traits and situational contexts, exhibits varied interaction dynamics, including metrics such as utterance frequency, silence episodes, attempted utterances, post-silence utterance attempts, and successful post-silence utterances. Furthermore, the table documents the utilisation frequency of different cognitive mechanisms (memory retrieval, goal summary, topic analysis, time analysis, and speech count statistics).

REFERENCES

- [1] N. Karanikolas, E. Manga, N. Samaridi, E. Tousidou, and M. Vassilakopoulos, "Large language models versus natural language understanding and generation," in *Proceedings of the 27th Pan-Hellenic Conference on Progress in Computing and Informatics*, pp. 278–290, 2023.
- [2] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, et al., "Improving language understanding by generative pre-training," *Technical report, OpenAI* 0, 2018.
- [3] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al., "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774* 0, 2023.
- [4] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al., "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971* 0, 2023.
- [5] 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.
- [6] 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.
- [7] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann, et al., "Palm: Scaling language modeling with pathways," *Journal of Machine Learning Research* 24(240), pp. 1–113, 2023.
- [8] D. H. Hagos, R. Battle, and D. B. Rawat, "Recent advances in generative ai and large language models: Current status, challenges, and perspectives," *IEEE transactions on artificial intelligence* 0, 2024.
- [9] 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.
- [10] Z. Xi, W. Chen, X. Guo, W. He, Y. Ding, B. Hong, M. Zhang, J. Wang, S. Jin, E. Zhou, et al., "The rise and potential of large language model based agents: A survey," *Science China Information Sciences* 68(2), p. 121101, 2025.
- [11] Y. Lei, T. Wang, J. Lian, Z. Hu, D. Lian, and X. Xie, "Humanllm: Towards personalized understanding and simulation of human nature," *arXiv preprint arXiv:2601.15793* , 2026.
- [12] F. Shareef, "Enhancing conversational ai with llms for customer support automation," in *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, pp. 239–244, IEEE, 2024.
- [13] 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.
- [14] J. Jiang, F. Wang, J. Shen, S. Kim, and S. Kim, "A survey on large language models for code generation," *ACM Transactions on Software Engineering and Methodology* 35(2), pp. 1–72, 2026.
- [15] C. Zhai, "Large language models and future of information retrieval: opportunities and challenges," in *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*, pp. 481–490, 2024.
- [16] H. Nachimovsky, M. Tennenholtz, and O. Kurland, "A multi-agent perspective on modern information retrieval," *arXiv preprint arXiv:2502.14796* 0, 2025.
- [17] Y. Li, C. Qian, Y. Xia, R. Shi, Y. Dang, Z. Xie, Z. You, W. Chen, C. Yang, W. Liu, et al., "Cross-task experiential learning on llm-based multi-agent collaboration," *arXiv preprint arXiv:2505.23187* 0, 2025.
- [18] R. Vasudev, M. Russak, D. Bikel, and W. Alshikh, "Accurate failure prediction in agents does not imply effective failure prevention," *arXiv preprint arXiv:2602.03338* , 2026.
- [19] 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.
- [20] S. S. Chowa, R. Alvi, S. S. Rahman, M. A. Rahman, M. A. K. Raiaan, M. R. Islam, M. Hussain, and S. Azam, "From language to action: a review of large language models as autonomous agents and tool users," *Artificial Intelligence Review* , 2026.
- [21] Z. Gao, J. Ma, X. Li, P. Li, N. Qu, Y. Wu, H. Liu, and J. Liu, "Aero: Autonomous evolutionary reasoning optimization via endogenous dual-loop feedback," *arXiv preprint arXiv:2602.03084* , 2026.
- [22] S. Hong, X. Zheng, J. Chen, Y. Cheng, J. Wang, C. Zhang, Z. Wang, S. K. S. Yau, Z. Lin, L. Zhou, et al., "Metagpt: Meta programming for multi-agent collaborative framework," *arXiv preprint arXiv:2308.00352* 3(4), p. 6, 2023.
- [23] Y. Wang, J. Chen, and H. Xiao, "Role-playing agents driven by large language models: Current status, challenges, and future trends," *arXiv preprint arXiv:2601.10122* , 2026.
- [24] 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.

- [25] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems* **33**, pp. 1877–1901, 2020.
- [26] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog* **1**(8), p. 9, 2019.
- [27] E. La Malfa, G. La Malfa, S. Marro, J. M. Zhang, E. Black, M. Luck, P. Torr, and M. Wooldridge, "Large language models miss the multi-agent mark," *arXiv preprint arXiv:2505.21298* **0**, 2025.
- [28] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, *et al.*, "A survey on large language model based autonomous agents," *Frontiers of Computer Science* **18**(6), p. 186345, 2024.
- [29] S. Zhang, J. Wang, R. Zhou, J. Liao, Y. Feng, Z. Li, Y. Zheng, W. Zhang, Y. Wen, Z. Li, *et al.*, "Memrl: Self-evolving agents via runtime reinforcement learning on episodic memory," *arXiv preprint arXiv:2601.03192* , 2026.
- [30] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafraan, K. Narasimhan, and Y. Cao, "React: Synergizing reasoning and acting in language models," in *International Conference on Learning Representations (ICLR)*, 2023.
- [31] Z. Xiu, D. Q. Sun, K. Cheng, M. Patel, Y. Zhang, J. Lu, O. Attia, R. Vemulapalli, O. Tuzel, M. Cao, *et al.*, "Astra-bench: Evaluating tool-use agent reasoning and action planning with personal user context," *arXiv preprint arXiv:2603.01357* , 2026.
- [32] M. Zhang, J. Kim, S. Xiang, J. Gao, and C. Cao, "Dynamic role assignment for multi-agent debate," *arXiv preprint arXiv:2601.17152* , 2026.
- [33] 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.
- [34] R. Takata, A. Masumori, and T. Ikegami, "Spontaneous emergence of agent individuality through social interactions in llm-based communities," *arXiv preprint arXiv:2411.03252* **0**, 2024.
- [35] C. Mu, H. Guo, Y. Chen, C. Shen, D. Hu, S. Hu, and Z. Wang, "Multi-agent, human-agent and beyond: a survey on cooperation in social dilemmas," *Neurocomputing* **610**, p. 128514, 2024.
- [36] 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.
- [37] C.-M. Chan, W. Chen, Y. Su, J. Yu, W. Xue, S. Zhang, J. Fu, and Z. Liu, "Chateval: Towards better llm-based evaluators through multi-agent debate," *arXiv preprint arXiv:2308.07201* **0**, 2023.
- [38] K.-T. Tran, D. Dao, M.-D. Nguyen, Q.-V. Pham, B. O'Sullivan, and H. D. Nguyen, "Multi-agent collaboration mechanisms: A survey of llms," *arXiv preprint arXiv:2501.06322* **0**, 2025.
- [39] Amazon Web Services, "Agent squad: Flexible and powerful framework for managing multiple ai agents and handling complex conversations." <https://github.com/awsllabs/agent-squad>, 2024. GitHub repository, Apache-2.0 License.
- [40] 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," *arXiv preprint arXiv:2307.07924* **0**, 2023.
- [41] Q. Wu, G. Bansal, J. Zhang, Y. Wu, B. Li, E. Zhu, L. Jiang, X. Zhang, S. Zhang, J. Liu, *et al.*, "Autogen: Enabling next-gen llm applications via multi-agent conversations," in *First Conference on Language Modeling*, 2024.
- [42] N. Spivack, S. Douglas, M. Cramers, and T. Connors, "Cognition is all you need—the next layer of ai above large language models," *arXiv preprint arXiv:2403.02164* **0**, 2024.
- [43] Z. Gu, X. Zhu, H. Guo, L. Zhang, Y. Cai, H. Shen, J. Chen, Z. Ye, Y. Dai, Y. Gao, *et al.*, "Agent group chat: An interactive group chat simulacra for better eliciting collective emergent behavior," *CoRR* **0**, 2024.
- [44] X. Chen, H. Yi, M. You, W. Liu, L. Wang, H. Li, X. Zhang, Y. Guo, L. Fan, G. Chen, *et al.*, "Enhancing diagnostic capability with multi-agents conversational large language models," *NPJ digital medicine* **8**(1), p. 159, 2025.
- [45] H. Li, Y. Q. Chong, S. Stepputtis, J. Campbell, D. Hughes, M. Lewis, and K. Sycara, "Theory of mind for multi-agent collaboration via large language models," *arXiv preprint arXiv:2310.10701* **0**, 2023.
- [46] O. P. John, E. M. Donahue, and R. L. Kentle, "Big five inventory," *Journal of personality and social psychology* **0**, 1991.
- [47] H. Jiang, X. Zhang, X. Cao, C. Breazeal, D. Roy, and J. Kabbara, "Personallm: Investigating the ability of large language models to express personality traits," in *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 3605–3627, 2024.
- [48] L. Cao, "Diaggpt: An llm-based and multi-agent dialogue system with automatic topic management for flexible task-oriented dialogue," *arXiv preprint arXiv:2308.08043* **0**, 2023.

A. Mechanism Specifications

In this framework, we implement six cognitive modules inspired by human reasoning: memory system, topic management, planning, temporal control, and utterance frequency tracking. These mechanisms simulate human thought processes, providing decision-support tools that the agent autonomously selects to activate. Upon triggering, each module returns structured data to inform subsequent decision-making, with the mechanism selection prompt illustrated in Fig. 6.

The **memory system**, adapted from generative agent [9] architectures, incorporates both storage and retrieval functions. During encoding, the agent first evaluates input significance through a perceptual module, assigning corresponding weight scores. The retrieval algorithm employs a multi-dimensional approach, assessing content relevance, temporal recency, and importance ratings.

The **topic management** mechanism, inspired by DiagGPT [48], continuously analyses dialogue history to identify active topics while maintaining a record of discussed subjects. Operating on three dimensions - conversational context, role-specific interests, and topic state - it guides the agent in maintaining, shifting, or pausing discussion threads.

The **planning module** adopts a hierarchical task decomposition strategy, breaking macro-objectives into executable sub-goal sequences. Initialisation involves generating predefined sub-goals based on scenario settings, with completion status markers. During interaction, the system dynamically evaluates progress, updating sub-goal completion states and their contribution towards overarching objectives.

The **temporal control** mechanism employs multi-parameter modelling to optimise utterance strategies dynamically. Input variables include dialogue start/end times and personal schedules. By computing available temporal resources against target progress, it generates situational reports to guide speaking rhythm adjustments.

The **utterance frequency tracker** quantitatively monitors participant speaking turns, constructing a statistical dictionary. This dataset provides objective reference for analysing participation patterns, helping identify dominant speakers and silent participants to optimise turn-taking strategies.

All modules follow modular design principles, maintaining functional independence through plugin-based implementation. This architecture supports extensibility, allowing integration of additional behavioural modules (e.g., predefined mention-response systems) triggered under specific conditions. The design significantly enhances behavioural anthropomorphism in artificial agents.

B. Speaking Mechanism Implementation Details

This section provides the complete technical specification of the dynamic speaking mechanism that implements the Self-Driven Order paradigm described in Section III-B. The mechanism operates in two sequential stages per dialogue round and incorporates adaptive temporal adjustments to simulate natural human turn-taking behavior.

1) *Willingness Screening*: Detailed formulation of the composite willingness score W_a for an agent a at turn t :

$$W_a(t) = \alpha \cdot S_{\text{topic}} + \beta \cdot S_{\text{goal}} + \gamma \cdot S_{\text{emotion}} + \delta \cdot S_{\text{personality}} \quad (1)$$

- S_{topic} : Topic relevance, computed as cosine similarity between the agent’s knowledge vector and the current dialogue embedding.
- S_{goal} : Goal urgency, measuring advancement toward the agent’s role-specific objective.
- S_{emotion} : Emotional expression need, derived from the agent’s affective state.
- $S_{\text{personality}}$: Personality-driven speaking propensity (e.g., extraversion factor).

The weights $\alpha, \beta, \gamma, \delta$ are role-specific parameters (typical range: [0, 1]). An agent proceeds to the temporal competition stage only if $W_a(t) \geq \theta$, where $\theta = 0.5$ is the default willingness threshold.

2) *Temporal Distribution and Sampling*: Agents that pass willingness screening are assigned a response delay T_i sampled from a personality-specific log-normal distribution:

$$T_i \sim \ln \mathcal{N}(\mu_a, \sigma_a^2) \quad (2)$$

Parameters μ_a (log-mean) and σ_a (log-standard deviation) are calibrated to agent personality:

- Proactive roles: $\mu_a = 0.3, \sigma_a = 0.2$
- Cautious roles: $\mu_a = 1.0, \sigma_a = 0.4$
- Neutral roles: $\mu_a = 0.6, \sigma_a = 0.3$

The agent with minimal T_i wins the speaking turn for that round.

3) *Dynamic Time Adjustment Mechanism*: To simulate cognitive continuity and prevent perpetual exclusion, response times are dynamically adjusted based on thought-content similarity. For each new thought T_{new} , we compute its maximum weighted similarity against historical thoughts:

$$S_{\text{max}} = \max_i (\text{sim}(T_{\text{new}}, T_{\text{old},i}) \cdot w_i) \quad (3)$$

where weights $w_i = \frac{i+1}{n}$ prioritize recent thoughts, and $\text{sim}(\cdot, \cdot)$ denotes cosine similarity. Using the average similarity \bar{S} of high-similarity items ($S > 0.7$), we adjust the distribution parameters:

$$\mu = \mu_0 e^{-k_1 \bar{S}} \quad (4)$$

$$\sigma = \sigma_0 e^{-k_2 \bar{S}} \quad (5)$$

where μ_0, σ_0 are baseline parameters, and $k_1 = 0.5, k_2 = 0.3$ are sensitivity coefficients. The adjusted response time is sampled as:

$$t \sim \text{LogNormal} \left(\ln(\mu), \sqrt{\ln \left(1 + \frac{\sigma^2}{\mu^2} \right)} \right) \quad (6)$$

This ensures faster responses when thoughts align with cognitive history (simulating priming) while maintaining natural variability.

4) Edge Case Handling and Parameters:

- **Tie-breaking:** If multiple agents share the minimal T_i , one is randomly selected.
- **No qualified agents:** If $W_a(t) < \theta$ for all agents, the system waits $\Delta t = 1.5s$ before re-evaluating.
- **Persistence factor:** For agents with unexpressed high willingness in consecutive turns, subsequent T_i values are multiplied by $\alpha = 0.7$.

TABLE III
KEY MECHANISM PARAMETERS AND TYPICAL VALUES.

Parameter	Description	Typical Value/Range
θ	Willingness threshold	0.5
μ_0	Baseline mean latency	2.0s
σ_0	Baseline std. deviation	1.0s
k_1, k_2	Similarity sensitivity coefficients	0.5, 0.3
$S_{\text{threshold}}$	Similarity cutoff	0.7
α	Persistence discount factor	0.7
Δt	Re-evaluation delay	1.5s

This integrated mechanism enables agents to exhibit human-like conversational dynamics through context-aware willingness assessment, personality-modulated response timing, and cognitive continuity modeling.

C. Experimental Settings

We employed GPT-4o as the base model. The global chat session was set to a maximum of 100 rounds or 30 minutes, terminating when either condition was met. For time simulation, we accounted for thinking time, speaking time, and silence duration. The sampling interval for speech was set to 10 seconds—if no speech occurred within 10 seconds, it was recorded as 10 seconds of silence. If speech occurred, the thinking and speaking times were aggregated to advance the timeline.

Each simulated role was assigned a predefined thinking time distribution, with adjustment parameters set to $k_1 = 0.5$ and $k_2 = 0.3$. The similarity threshold for adjustments was $threshold = 1.0$. Each scenario involved five roles, and we conducted five simulation runs with identical character settings to ensure reliable results.

Across multiple simulations, we implemented varied configurations and recorded extensive dialogue data. Although large language models exhibit inherent output variability, repeated experiments revealed consistent patterns. To highlight key phenomena and facilitate analysis, we selected representative cases. These examples reflect typical agent behaviours in complex scenarios, representing common observations across repeated simulations.

D. Theoretical Basis of the Binary Big Five Personality Model

The Big Five Personality Model is a classic and widely used framework in psychology for describing individual personality traits. It comprises the following five key dimensions:

- **Extraversion:** Measures an individual’s sociability and energy levels in social settings. This dimension contrasts extroverted with introverted traits.

- **Agreeableness:** Reflects an individual’s cooperativeness and friendliness in interpersonal interactions. This dimension contrasts agreeable with antagonistic traits.
- **Conscientiousness:** Assesses an individual’s self-discipline, organisation, and dependability. This dimension contrasts conscientious with unconscientious traits.
- **Neuroticism:** Indicates an individual’s emotional stability. This dimension contrasts neurotic with emotionally stable traits.
- **Openness to Experience:** Describes an individual’s receptiveness to new ideas and creativity. This dimension contrasts open with closed traits.

When constructing a character’s personality, we randomly select relevant trait adjectives from these five binary dimensions to form a foundational personality framework. Subsequently, leveraging the capabilities of large language models, we expand this framework into 10 detailed and precise descriptions. These descriptions closely align with the selected trait adjectives, comprehensively illustrating the character’s thought processes and conversational style. The prompt is shown in Fig. 7.

Finally, these 10 descriptions serve as the core reference for the character’s personality, guiding and constraining their behaviour in subsequent simulated tasks to ensure consistency with the established traits.

E. Personality Trait Assessment Using the Big Five Inventory

In this study, we employed the Big Five Inventory (BFI) [46] to assess individual personality traits. The BFI is a widely utilised self-report questionnaire specifically designed to measure the Big Five personality traits, namely Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. These traits provide a comprehensive reflection of individual differences in social behaviour, emotional responses, and cognitive styles.

- 1) **Data Collection:** The experimental conversations were segmented into groups of 30 turns each (for instance, a 90-turn conversation was divided into three groups). Each group of conversation served as an independent input for GPT-4o to evaluate. This segmentation method was adopted to simulate the continuity and coherence of real-life conversational scenarios while facilitating detailed analysis of each group of dialogue.
- 2) **BFI Scoring:** Based on the 44-item BFI questionnaire, GPT-4o rated the personality traits manifested in each group of dialogue on a scale from 1 to 5, where 1 indicates strong disagreement and 5 indicates strong agreement. This rating method allows for the quantification of personality traits exhibited in the dialogue, facilitating subsequent statistical analysis.
- 3) **Data Integration and Analysis:** The ratings from all scenarios were categorised and integrated according to the 10 specific facets of the Big Five personality traits. Specifically, the ratings for each facet were aggregated to calculate the mean and confidence intervals for each

trait. This analytical approach enables a more comprehensive understanding of the personality traits exhibited in the dialogue and allows for the assessment of their consistency and stability.

F. Statistical Analysis

Tab. II provides a detailed statistical analysis of the behavioral patterns and mechanism selection strategies of the five autonomous agents during a 30-minute simulated group discussion. Each agent, governed by distinct personality traits and situational contexts, exhibits varied interaction dynamics. The table includes the following data for each agent:

- **Character:** Agent identifier and role
- **Personality Traits:** Big Five personality dimensions configuration
- **Special Settings:** Unique scenario constraints or objectives
- **Spoken:** Total number of successful utterances
- **Wanting to Speak:** Total attempts to speak
- **Silent:** Number of silence episodes
- **Mechanism (On-Off):** Activation frequency for each cognitive mechanism (Memory Retrieval, Goal Summary, Topic Analysis, Time Analysis, Speech Count Statistics)
- **Wanting to Speak After Silence:** Attempts to speak following group silence
- **Speaking After Silence:** Successful utterances after group silence

G. Simulated Cases

We present some examples of special behaviors in Fig. 5, including the natural emergence of leadership, conflict and mediation, as well as innovative proposals. A comprehensive analysis is provided in Section Experiments—Special Behaviour of the main text.

PROMPT:

As [{name}], you possess human-like cognitive abilities that enable you to determine how and what to think during conversations. Below is a list of commonly used human thinking mechanisms. Your current task is to decide whether to activate these functions based on the received information and the context of the ongoing dialogue.

Available mechanisms:

- **Memory Retrieval**: Used when recalling specific information; retrieves relevant details from memory. Requires specifying the retrieval content.
- **Goal Summary**: Used when discussing objectives; evaluates the progress toward goals based on historical dialogue.
- **Topic Analysis**: Used to summarize discussion themes; analyzes dialogue history to identify covered topics. Note: Only provides summaries of past discussions.
- **Time Analysis**: Analyzes the current time and personal schedule to provide insights about temporal context.
- **Speech Frequency Statistics**: Summarizes the speaking frequency of participants based on historical dialogue.

Follow these steps for decision-making:

1. Read the mechanism descriptions to understand their purposes and characteristics.
2. Review the received information to grasp the dialogue context and current content.
3. Analyze the character's personality, current emotions, social relationships, and other factors to understand their thinking style and objectives.
4. Based on the received information and dialogue context, determine whether [{name}] should activate these mechanisms.

Examples

... ..

Current Context

[{situation_summary}]

Current Dialogue:

[{dialogue}]

Mechanism Usage Statistics

- Total dialogue rounds: [{dialogue_round}] (One round is counted when someone speaks or after 10 seconds of silence).
- Mechanism usage records: [{mechanism_stats}]

Note: 'last_used' indicates the last round the mechanism was activated, and 'count' shows the total usage frequency. Activate mechanisms only when necessary.

Important Notes:

- Prioritize strict adherence to character traits in your reasoning.
- Focus on analyzing the last utterance in the dialogue; other content is for reference only.
- **Alignment with the character's personality** is paramount—avoid excessive speculation.
- When referring to the character you are roleplaying, use the first-person "I" instead of the third person.
- Output format must be a JSON dictionary without additional symbols (refer to the example format):

```
{
  "thinking_process": {
    "key_info": "Key dialogue information",
    "character_analysis": "Character trait analysis",
    "decision_reason": "Reason for mechanism selection, must align with the 'decisions' section"
  },
  "decisions": {
    "memory_retrieval": {
      "active": true/false,
      "reason": "Reason for activation/deactivation"
    },
    "goal_summary": {
      "active": true/false,
      "reason": "Reason for activation/deactivation"
    },
    "topic_analysis": {
      "active": true/false,
      "reason": "Reason for activation/deactivation"
    },
    "time_analysis": {
      "active": true/false,
      "reason": "Reason for activation/deactivation"
    },
    "speech_frequency": {
      "active": true/false,
      "reason": "Reason for activation/deactivation"
    }
  }
}
```

Fig. 6. Illustration of the Prompt for Agent Mechanism Selection Decision-Making. The text in square brackets represents input variables; some examples have been omitted due to space constraints.

PROMPT:

You are now a personality trait analysis expert. Based on the following Big Five personality trait combination: [{personality}]

Please generate 10 specific descriptions of thinking patterns and conversational characteristics based on the above trait combination. Requirements:

Output format:

- Each description begins with a number (1-10) and contains 10-30 words
- All descriptions are output consecutively, with only one line break after each
- No line break after the last description

Descriptions must focus on:

- Thinking patterns: How problems are approached, decision-making processes, information processing
- Conversational characteristics: Expression style, response tendencies, interaction preferences

Example output (based on a different trait combination):

1. During discussions, tends to express intuitive ideas quickly (extraversion) but adjusts tone promptly to avoid offending others (agreeableness)
2. When analysing problems, often anticipates worst-case scenarios (neuroticism) while actively seeking innovative solutions (high openness)

Ensure the descriptions:

- Reflect the combined effects of multiple traits logically, with consistent and harmonious expressions
- Are concrete and observable in conversations
- Avoid referencing physical actions, focusing solely on cognition and verbal expression
- Exclude contradictory behavioural descriptions

INPUT:

['introverted', 'antagonistic', 'unconscientious', 'neurotic', 'closed to experience']

OUTPUT:

1. Prefers solitary reflection on problems, often ruminating on negative outcomes before making decisions, leading to a cautious approach in conversations.
2. Engages in discussions with a critical and sometimes confrontational tone, often dismissing alternative viewpoints without much exploration of their validity.
3. Tends to focus on the flaws in others' arguments, often undervaluing constructive feedback, which can stifle collaborative conversations.
4. In decision-making, often weighs potential threats heavily, leading to indecisiveness, especially when faced with unfamiliar situations or new ideas.
5. Communicates in a straightforward manner, often prioritizing blunt honesty over tact, which can result in misunderstandings or hurt feelings in social interactions.
6. Shows a tendency to overanalyze social cues, often misinterpreting neutral statements as personal attacks, reflecting high levels of neuroticism in interactions.
7. Prefers to rely on established routines and familiar topics during conversations, avoiding deep dives into new concepts or experiences that might cause discomfort.
8. When confronted with conflicting opinions, tends to become defensive rather than open to dialogue, reinforcing a stance of antagonism in discussions.
9. Often engages in negative self-talk when facing challenges, which can influence conversational dynamics by projecting insecurity or pessimism onto others.
10. Displays a reluctance to share personal experiences or emotions, leading to superficial exchanges that lack depth and emotional connection.

Fig. 7. Character personality generation prompt and examples.