基于大语言模型的多智能体系统：博弈论视角

Jianing Hao[1] , **Han Ding**[2] , **Yuanjian Xu**[1] , **Tianze Sun**[3] , **Ran Chen**[4] , **Wanbo Zhang**[5] , **Guang Zhang**[1,*], **Siguang Li**[1,*]

[1]香港科技大学（广州）， [2]北京航空航天大学[3]哈尔滨工业大学， [4]OpenCSG， [5]复旦大学，

jhao768@connect.hkust-gz.edu.cn，handing@buaa.edu.cn，

yxu085@connect.hkust-gz.edu.cn，suntianze@stu.hit.edu.cn，schen@opencsg.com，

23302010062@m.fudan.edu.cn，guangzhang@hkust-gz.edu.cn，siguangli@hkust-gz.edu.cn

arXiv:2601.15047v1 [cs.MA] 21 Jan 2026

摘要

大语言模型（LLM）展现出了强大的推理、规划和沟通能力，使其能够作为自主智能体在开放环境中运行。单智能体系统在适应性和协调能力方面仍存在局限，而近期的发展使研究重点转向由相互作用的大语言模型组成、追求合作、竞争或混合目标的多智能体系统（MAS）。这一新兴范式为研究智能体之间的社会动态和策略行为提供了强大的试验平台。然而，当前研究仍较为零散，缺乏统一的理论基础。为弥补这一空白，我们从博弈论视角对基于大语言模型的多智能体系统进行了全面综述。我们围绕博弈论的四个关键要素——参与者、策略、收益和信息——对现有研究进行梳理，建立了一个系统框架，以帮助理解和比较相关研究，并指导未来对基于大语言模型的多智能体系统的设计与分析。

1 引言

多智能体系统（MAS）长期以来一直是研究自主实体之间交互、决策和战略行为的基础框架。经典研究通过博弈论模型 [纳什、 1950], 分布式优化 [Sandholm、 1999], 强化学习 [Littman,1994], 以及基于智能体的建模 [Epstein, 1999], , 提供了用于理解协调、竞争、谈判和涌现集体行为的形式化工具。

大语言模型（LLM）的最新进展催生了一种新型多智能体系统范式，其中智能体能够进行推理 [Wei等，规划 [Huang 等，并运用自然语言进行交流 [Li 等。在这类系统中，多个基于大语言模型的智能体在复杂且开放式的环境中互动，展现出丰富多样的行为，包括合作 [Chen 等、竞争 [Zhu等、谈判 [Hua 等，社会学习，以及其他涌现现象 [Ashery 等， 2025]。

延续对战略决策系统的研究脉络 [Silver 等人， 2016], 先前研究在多种情境中探索了基于大语言模型的多智能体系统，包括任务分解与规划 [Hong 等人，2024],不确定性下的决策 [Agashe 等人， 2025],多智能体协调 [Wu 等人， 2024], 社会模拟与行为建模 [Park 等人，2023], 多方辩论或对话 [Qian 等人，2024], 以及竞争性策略游戏 [Zhu 等人，2025]。这些研究凸显了语言模型多智能体系统在洞察智能体交互、战略推理、通信协议和涌现集体行为方面的潜力，但目前仍缺乏一种能够系统梳理这些进展的统一理论视角。

多智能体系统中的每一次交互，背后都存在一个策略决策问题，而这正是博弈论旨在建模的领域。随着由大语言模型驱动的智能体能力不断增强、自主性不断提高，通过博弈论视角理解其行为变得至关重要。本综述采用**博弈论视角**，系统梳理关于大语言模型驱动的多智能体系统的现有研究。本文的贡献主要有以下三点:

- 我们提出了一种新颖的博弈论框架，通过四个核心要素——参与者、策略、收益和信息——对基于大语言模型的多智能体系统进行分类。这一统一视角有助于将经典博弈论与现代大语言模型驱动的研究相结合，从系统层面分析智能体交互。
- 我们系统回顾了当前关于基于大语言模型的多智能体系统的研究，揭示了现有研究的见解与局限。
- 我们指出了关键研究空白，并提出了具有前瞻性的研究方向，重点关注优化均衡协调、设计激励相容的通信协议，以及部分可观测条件下的信息结构建模。

2 背景与分类体系

2.1 博弈论基础

博弈论是对策略决策的数学研究，其中多个智能体彼此交互，每个智能体的结果取决于其他智能体所做的选择。它研究的是

*通讯作者。



图1：基于大语言模型的多智能体系统的博弈论框架，展示了博弈四个核心要素——参与者、策略、收益和信息——之间的动态交互。

理性决策者之间的策略互动，其中每位参与者都力求最大化自身收益。标准型博弈由元组 $\Gamma = (N, (S_i)_{i\in N}, (u_i)_{i\in N})$ 定义，其中 $N = \{1, \ldots, n\}$ 是参与者的有限集合，S_i 是参与者 i 可选的纯策略有限集合，而 $u_i: \prod_{j\in N} S_j \to \mathbb{R}$ 是参与者 i 的收益函数，为每个联合策略组合分配一个实数值奖励。策略组合是一个向量 $\mathbf{s} = (s_1, \ldots\ldots, s_n) \in S_1 \times \cdots S_j \cdots \times S_n$，表示所有参与者的策略选择。在此背景下，我们假设参与者都是理性的，并通过选择能够优化自身收益的策略，依据其可获得的信息来最大化期望效用。

策略可以是纯策略，也可以是混合策略。参与者 i 的纯策略是一个元素 $s_i \in S_i$，而混合策略是在 S_i 上的概率分布 $\sigma_i \in \Delta(S_i)$。这里，$\Delta(S_i)$ 表示 S_i 上概率测度构成的单纯形。混合策略组合 $\sigma = (\sigma_1, \ldots, \sigma_n)$ 为每位参与者的动作指定一个分布。在策略组合 σ 下，每个纯策略组合 $s = (s_1, \ldots, s_n)$ 以概率 $\prod_{j=1}^{n} \sigma_j(s_j)$ 被选中，参与者 i 的期望收益为 $U_i(\sigma) = \sum_{s\in S} u_i(s) \prod_{j=1}^{n} \sigma_j(s_j)$。纯策略 s_i 可以视为一种特殊的混合策略，其概率集中于单个行动。

策略选择既受收益影响，也受信息影响。在完全信息下，所有收益函数和策略空间 S_i 都是共同知识，参与者可以准确预测他人的行为。相比之下，在不完全信息下，每个参与者可能拥有私人知识或信念，将其建模为 θ_i。无论信息结构如何，参与者都会根据自己掌握的信息选择能够使期望效用最大化的策略。策略的吸引力取决于其对应的收益，而可用信息的结构决定了策略如何受条件约束，以及信念如何更新。当任何参与者都无法通过单方面改变策略来改善自身结果——假设其他参与者保持策略不变时，就达到了纳什均衡。这一概念对于多智能体系统至关重要，因为其中的智能体会在竞争或合作环境中寻求稳定的策略。

2.2 多智能体系统作为博弈论环境

多智能体系统由多个自主智能体组成，它们在共享环境中进行交互。与单智能体系统不同，多智能体系统利用分布式智能，使智能体能够分工、响应局部信号，并通过协作或竞争解决复杂问题。传统的多智能体系统依赖基于规则的逻辑和局部策略，但在局部指标可能无法准确反映全局行为的情况下，仍面临挑战；在博弈中的学习动态也存在类似问题 [Fudenberg and Levine, 1998]。然而，随着大语言模型的发展，智能体能够进行自然语言推理、灵活的任务协商和战略协调。从博弈论视角来看，每个智能体都被建模为一名参与者 $i \in N$，拥有自己的策略空间和效用函数。Shoham and Leyton-Brown [Shoham and Leyton-Brown, 2008] 将多智能体系统描述为由自主实体构成的系统，这些实体可能具有不同的利益或信息；这与博弈论中参与者拥有不同收益或私人信息的概念相呼应。因此，任何多智能体系统交互都可以表示为标准式博弈：智能体可能采取的行动对应于参与者的策略，而智能体的目标则由其针对联合行动组合定义的收益函数编码。

通信与协调是多智能体系统的关键特征。通信可以建模为博弈前信号传递：智能体在选择行动之前交换消息或共享观测。这类信号使智能体能够共享私人信息或协调预期。当智能体拥有共同目标时，就会出现协调问题：如果它们的收益一致，多智能体系统本质上就是在进行合作博弈。更一般地说，协调自身利益至上的智能体，既需要在存在多个一致策略组合的博弈中寻找均衡，也需要解决机制设计问题，使个人激励与集体目标保持一致，尤其是在通信可能带有策略意图或涉及虚假陈述的情形下 [Kamenica and Gentzkow,2011]。多智能体系统中的自主性指不存在中央控制器。每个智能体都根据自身收益和可用信息独立作出决策。这体现了非合作博弈中分散决策的假设。智能体可能掌握有关环境或其他参与者偏好的私人信息或不完全信息，这类似于贝叶斯博弈中参与者的类型。总之，多智能体系统的特征可直接对应到博弈论概念：通信对应博弈中的信号传递或消息传递，协调体现为博弈中的信号传递，协调对应均衡选择或联盟形成，而自主性则体现为独立且效用最大化的参与者。

基于大语言模型的多智能体系统具有独特的特征。在这些系统中，智能体由大语言模型驱动，并通过自然语言进行交互。有效的行动空间由潜在的话语或文本序列构成，具有组合性且没有固定边界。通信本质上由语言驱动，能够实现丰富的谈判与解释。近期研究表明，大语言模型智能体能够自发形成

通信协议和角色。通过对话，这些智能体能够协商资源、委派任务，或在合作与竞争之间取得平衡，展现出合作、竞争、谈判和社会互动等涌现行为。这种在激励机制下进行策略性通信的能力，明确地将大语言模型之间的交互与经典信号理论联系起来。这些特征表明，基于大语言模型的智能体作为博弈论参与者，运用表达力极强、由语言驱动的策略，并动态交换信息，因此其交互比传统多智能体系统模型更复杂、更多变。

2.3 基于博弈论要素的分类体系

在多智能体系统中，智能体并非孤立运作；它们彼此交互、影响并回应。博弈论提供了一个形式化框架，通过将智能体视为选择相互依存的策略性参与者，对这些交互进行建模。我们的分析围绕博弈论的四个核心要素展开：参与者、策略、收益和信息。这些要素构成了任何策略模型的基础组成部分。我们聚焦这四个维度，因为它们共同界定了智能体的身份、可采取的行动、重视的事物以及掌握的信息。值得注意的是，我们不将行动视为独立于策略的要素，因为行动只是具体的策略选择。此外，我们不将均衡视为基本要素，因为它是基于其他四个组成部分推导出的结果。在标准型博弈中，参与者、其策略集及其收益函数完全决定了策略交互。我们将信息纳入一个独立维度，以体现智能体的私人信息或类型结构，并区分完全信息情形与不完全信息情形。

这些要素会动态地相互作用。每个参与者 i 都有一个策略空间 S_i 和收益 u_i，但具体选择哪些策略取决于 i可获得的信息。收益决定激励：在给定可用信息的情况下，参与者会选择使其效用最大化的策略。反过来，所选择的策略及其结果也可能揭示额外信息，从而影响后续的收益和决策。图1展示了这些相互依赖关系：参与者（P）根据收益激励和信息信号选择策略（S），而由此产生的策略组合会带来收益

以及潜在的新信息。

3 多智能体系统中的博弈论要素

3.1 参与者

在基于大语言模型的多智能体系统中，智能体（参与者）之间的交互远不止传统的双人博弈。根据其设计和目标，这些智能体在能力、目标或信息获取方面可能各不相同，从而形成异质群体，并产生多样的策略动态。从博弈论视角来看，这些交互大致可分为合作、竞争和混合动机系统，反映了智能体目标之间不同程度的一致性，如图2所示。

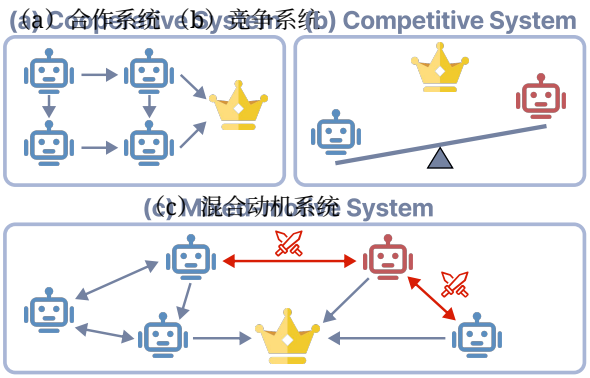


图2：大语言模型参与者的三种交互结构示意图。

合作系统。 合作系统中的智能体追求共同或一致的目标，通常通过社会福利函数 $W(a)$ 来形式化表示，该函数汇总了个体效用。

$$W(a) = f\big(u_1(a), \ldots, u_N(a)\big), \quad a^* = \arg\max_{a \in A} W(a),$$

其中，$A = \prod_i A_i$ 表示联合行动空间，f： $\mathbb{R}^N \to \mathbb{R}$是一个聚合映射。常见的选择是加权和 $W(a) = \sum_i w_i u_i(a)$（线性 f）；而复杂场景可能需要采用非线性聚合，以满足公平性或对风险敏感的目标。合作型大语言模型智能体实际上组成了一支乐观的团队，共享信息并协调行动，以找到能够最大化联合福利的行动组合 a^* 。

近期研究已证明，合作型大语言模型智能体在具有挑战性的任务 [上表现有效 Zhu 等， 2025]。例如，Chain-of-Agents (CoA) 框架 [Zhang等，2024c] 将长上下文问题拆分为多个片段，由顺序工作的“工作者”智能体处理，随后交由“管理者”智能体处理，从而能够对较长的输入进行复杂推理。COPPER [Bo 等， 2024] 系统使用自我反思型智能体，通过反事实奖励对其输出进行微调，以提高贡献质量。CAMEL [Li 等， 2023] 等基于角色的对话框架采用“助手—用户”角色扮演来模拟类人团队合作，促成涌现的协作行为。其他架构，如 MetaGPT [Hong 等， 2024], 则在软件工程任务中为智能体分配项目经理、架构师和开发者等专门角色；而 AutoGen [Wu 等， 2024] 则提供了一个用于编排多智能体交互的灵活框架。实证评估表明，语言媒介协调能够显著提升问答、编程和规划等任务的性能。

竞争系统。 在竞争系统中，智能体具有冲突目标。这些目标可以形式化为（可能是零和）博弈 $G = (\mathcal{N}, A_i, u_i)$，其中每个智能体$i$ 都会选择策略 $\sigma_i \in \Delta(A_i)$，以最大化其期望收益 $U_i(\sigma)$。纳什均衡 $(\sigma_1^*, \ldots\ldots, \sigma_N^*)$ 满足相互的最佳响应条件：

$$U_i(\sigma_i^*, \sigma_{-i}^*) \geq U_i(\sigma_i, \sigma_{-i}^*) \quad \forall i, \sigma_i.$$

对于双人零和博弈，这可简化为经典的极小极大均衡：

$$\max_{\sigma_1} \min_{\sigma_2} U_1(\sigma_1, \sigma_2) = \min_{\sigma_2} \max_{\sigma_1} U_1(\sigma_1, \sigma_2).$$

大语言模型智能体已在对抗性基准测试和博弈中接受评估。例如，GTBench [Duan等人， 2024]是一个博弈论基准测试，包含十种经典博弈（棋盘游戏、纸牌游戏、谈判、拍卖），用于检验纯粹的战略推理。另一个测试平台 GameBench [Costarelli等人， 2024], 涵盖九种多样的博弈环境，结果表明，GPT-4 在零和博弈中的表现仍难以达到人类水平。专用竞争框架（如多智能体辩论系统 [Cheng等人， 2024], ）让智能体就对立观点展开辩论，从而完善自身答案。实证研究表明，尽管在多步策略预判方面仍有局限，但此类竞争练习有助于大语言模型展现长期规划和对手建模能力。

混合动机系统。 大多数现实的多智能体环境都融合了合作与竞争。在混合动机博弈中，每个智能体都会权衡自身利益与集体福利，可表示为：

$$u_i(a) = \alpha_i \, v_i^{\text{self}}(a) + (1 - \alpha_i) \, v^{\text{col}}(a), \quad \alpha_i \in [0, 1],$$

其中， v_i^{self}是智能体 i的私人收益， v^{col} 是共享的团队收益。参数 α_i 表示自身利益与协作的权衡程度。在此类系统中设计激励通常涉及机制设计：例如，引入转移支付 $t_i(a)$，以奖励或惩罚智能体，从而使诚实且以团队为导向的行为成为一种均衡。一个常见要求是满足激励相容（IC）和预算平衡（BB），这与经典机制设计原则一致 [Mas-Colell等人， 1995]:

$$\text{IC: } \sigma_i^* \in \arg\max_{\sigma_i} \mathbb{E}[u_i(a) + t_i(a)], \quad \text{BB: } \sum_i \mathbb{E}[t_i(a)] = 0.$$

这些约束确保每个智能体无需外部补贴，遵循预期策略就是最优选择。

近期的研究开始探索混合动机的大语言模型系统。Orner 等人 [Orner 等人， 2025] 研究了混合动机博弈中智能体的解释方法，例如外交和迭代囚徒困境的变体，并强调合作与竞争交织如何造成复杂的社会困境。作为对这些实证研究的补充，Duetting 等人 [Duetting等人， 2024] 从机制设计的角度出发，将大语言模型视为生成内容拍卖中的竞价智能体；该拍卖经过正式设计，以确保激励相容性和如实报告。此外，MAC-SPGG 框架 [Liang 等人， 2025] 通过序贯公共物品博弈处理动机的策略对齐。该框架证明，通过精确调整公共物品奖励，可以诱导出子博弈完美纳什均衡（SPNE），从而促进普遍合作。总体而言，这些多样的方法——从拍卖理论到序贯博弈建模——表明，精心设计的激励措施能够有效对齐大语言模型的动机并维持稳定协作，即使个体目标彼此不同。

3.2 策略与均衡

有效的策略制定对于多智能体大语言模型系统至关重要，因为智能体往往需要兼顾合作与竞争。基于 CICERO [(FAIR)† 等人等语言驱动的策略智能体， 2022], LLM-Nash 框架 [Zhu， 2025] 将每个智能体建模为选择推理提示作为策略，并在提示空间上定义“推理均衡”。不同于参与者完全理性的经典博弈，LLM-Nash 通过明确建模推理过程来刻画有限理性。该框架表明，这些推理均衡可能偏离经典纳什结果，反映出大语言模型智能体独特的策略行为。这说明，大语言模型能够通过迭代对话和提示逐渐形成新的类均衡行为，从而在非合作环境中有效近似博弈论推理。

自博弈是策略形成的另一种强大工具。在自博弈中，智能体与自身的副本进行竞争，并通过试错不断改进策略 [Zhang 等，2024b]。这种方法已应用于多智能体强化学习（MARL）：智能体不仅通过与固定环境交互来学习，还会依据自身行为随时间推移不断改进策略。近期研究表明，自博弈能让智能体探索广泛的战术与反制策略，从而显著增强大语言模型的策略能力。通过强化学习（RL）和策略优化，大语言模型能够调整自身行为，在重复或不断演变的策略环境中取得更好的结果。这一迭代过程使大语言模型得以逐渐收敛至最优策略。SPIRAL [Liu 等， 2025] 借助这一思路，让两个大语言模型智能体彼此进行多轮零和语言博弈。它采用完全在线的多轮多智能体强化学习系统，使所有模型参数在自博弈过程中持续更新。SPIRAL 展示了竞争性自我博弈不仅能推动智能体形成强大的游戏策略（类似于达到均衡），还能赋予它们通用的问题解决战术。MARSHAL [Yuan 等， 2025] 框架同样表明，策略博弈中的自博弈能够泛化到多种推理任务。MARSHAL 在合作博弈和竞争博弈中采用回合级优势估计与自博弈，研究发现，这能增强大语言模型的策略能力，并提升其在多智能体基准测试中的表现。

3.3 收益

在多智能体系统中，收益结构在设计有效机制方面发挥着关键作用。每个智能体的目标都是通过某些策略最大化自身奖励，但在多智能体环境中，智能体的奖励往往并不一致。挑战在于设计适当的奖励机制，使个体目标与集体目标保持一致，确保智能体的行为既能优化个体成就，也能提升系统整体效率。

机制设计。 有效的机制设计要求智能体的激励结构与系统的集体目标保持一致。在多智能体环境中，设每个智能体 i 都有一个个体奖励函数 $R_i(s, a)$，而系统的整体目标由社会福利函数表示

$W(s,a) = f(R_1(s,a)\ldots,R_N(s,a))$，其中$f(\cdot)$可能代表公平性、效率或安全性等因素的综合考量，其概念与微观经济学中的福利最大化相一致 [Mas-Colell 等， 1995]。当最大化 R_i 未必能改善 W时，就会产生目标不一致，而这种不一致可能导致系统结果不理想。

为量化这种不对齐，我们定义联合策略 $\pi = (\pi_1, \ldots, \pi_N)$ 下每个智能体的偏离收益为

$$\Delta_i(\pi) = \max_{\pi_i'} \mathbb{E}[R_i(s, a_i', a_{-i})] - \mathbb{E}[R_i(s,a)]\,,$$

该值表示智能体 i 通过单方面偏离联合策略所能获得的收益。合作机制应在提升集体目标的同时，尽量降低这些偏离收益。通用对齐框架可表示为：

$$\begin{aligned} \max_{\mathcal{M}} \quad & \mathbb{E}[W(s,a)] \\ \text{s.t.} \quad & \mathbb{E}[R_i(s,a)] \geq \mathbb{E}[R_i(s,a_i',a_{-i})]\,, \quad \forall i,\ \forall \pi_i', \end{aligned}$$

其中， $\mathcal{M}$表示为确保激励相容性而设计的奖励或转移机制。该对齐机制确保智能体的行为受到集体利益的驱动，同时尽量减少自利偏离。

奖励塑形。 奖励塑形提供了一种实用方法，可在保留智能体局部自主性的同时调整其激励。塑形后的奖励可以采用以下形式：

$$R_i'(s,a,s') = R_i(s,a,s') + \gamma\,\Phi_i(s') - \Phi_i(s),$$

其中， $\Phi_i(s)$是一个势函数，可调整时间信用分配，而不改变合作最优解。通过这种方法，奖励塑形能够加快收敛，并维持目标各异的智能体之间的稳定性。例如，COPPER [Bo 等， 2024] 提出将自我反思作为促进合作的方法，智能体通过内部调整激励，更好地完成协作任务。虽然没有给出明确的奖励调整数学形式，但这种方法可以理解为引入新的激励机制，鼓励智能体反思自身对合作的贡献。结合机制设计，奖励塑形可确保智能体在保持个体自主性的同时，始终与系统级目标保持一致。这种方法尤其适用于智能体必须在个人激励与集体表现之间取得平衡的多智能体系统。

惩罚与监管机制。 除了奖励塑造之外，惩罚与监管机制也是调节智能体行为的关键工具。通过引入一种

惩罚函数 $\Psi_i(s,\ a)$ 可以抑制伦理违规或资源滥用等不良行为，进一步引导智能体采取有益于社会的行为。经过调节的奖励可表示为：

$$\tilde{R}_i(s,a,s') = R_i'(s,a,s') - \lambda_i\,\Psi_i(s,a),$$

其中 $\lambda_i > 0$控制监管执行的力度。这一表述将软约束直接融入学习信号，使智能体能够动态权衡性能与合规性。对于安全关键型或

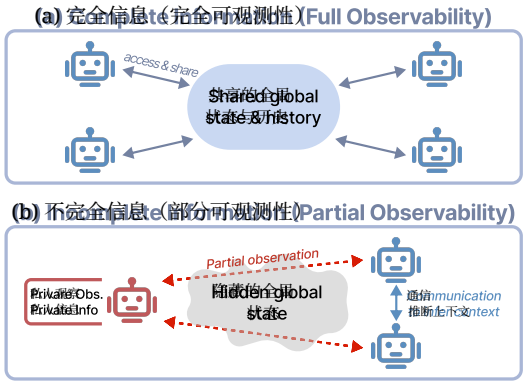


图3：基于大语言模型的参与者之间信息结构的示意图。

伦理敏感型系统，例如涉及人际互动或资源管理的系统，基于惩罚的规制对于确保智能体不会利用或损害系统至关重要。还可以在策略层面施加明确的监管约束，将违反安全性或伦理规范的行动排除在可行集合之外：

$$\pi_i(a|s) = 0 \quad \text{if} \quad \Psi_i(s,a) > \tau_i,$$

其中， τ_i 是一个阈值，超过该阈值的行动被视为不可接受。

近期研究 [Reid 等, 2025; Li 等, 2024; Zhang 等, 2024a] 强调了在多智能体系统中纳入监管信号的重要性，并表明整合显式或隐式的监管反馈——无论是语言形式、基于反馈的形式，还是自我惩罚——都能显著提升伦理合规性和长期系统稳定性。FinCon [Yu 等,2024], CORY [Ma 等， 2024], 以及 ProAgent [Zhang等, 2024a] 等实证框架展示了如何在实践中通过结构化奖励信号、语言反馈或反思性评估来落实此类对齐策略。

3.4 信息

系统的信息结构从根本上决定了智能体之间的策略复杂性和协调潜力。设 $\mathcal{I}_i(s)$ 表示状态 s下智能体 i 可获得的信息。此类系统的策略形式化可表示为一个贝叶斯博弈：

$$U_i(a_i, a_{-i}) = \mathbb{E}_{\theta_i \sim p(\theta_i|\mathcal{I}_i)}[R_i(s, a_i, a_{-i}, \theta_i)]\,,$$

其中， θ_i 表示私人信息或信念。这将大语言模型介导的交互与经典的不完全信息博弈联系起来。因此，均衡效率在很大程度上取决于 $\mathcal{I}_i$ 的结构以及其在智能体之间的共享方式。

完全可观测性。 在完全信息下，所有智能体共享全局状态，并且可以观察彼此的行动和收益（如图3 (a)所示）。这种透明性有助于联合规划，但需要更多的通信或计算。AgentVerse框架 [Chen 等， 2024] 展示了完整的信息共享如何提升专家智能体协作的性能。同样，Chain-of-Agents [Zhang等， 2024c] 和 COPPER [Bo等， 2024]框架利用完全可观测性来

促进流畅的序列推理和反思性适应。完全可观测性促进了基于角色的涌现式协作以及更强的泛化能力。即使是在 CAMEL [Li 等， 2023], 这类更简单的协作设置中，共享的完整状态也能使两个被分配角色的大语言模型紧密协调，并超越单智能体基线。从博弈论视角来看，完全信息近似于共同知识环境，在这种环境中，每个参与者对于其他参与者所掌握信息的信念趋于一致，从而简化均衡计算。

部分可观测性。 相比之下，不完全信息更贴近现实，但也更具挑战性。每个智能体都会接收一个信号或观测 θ_i，并必须根据一个信念分布 $b_i(s) = P(s \mid \theta_i)$ 选择动作（如图 3（b）所示）。在此类环境中，最优策略旨在最大化期望回报：

$$\pi_i^*(\theta_i) = \arg\max_{\pi_i} \mathbb{E}_{s\sim b_i,\, a_i\sim\pi_i(\theta_i)}[R_i(s, a_i, a_{-i})] \,.$$

Li 等人 [Li 等人， 2024] 提出了一种语言基础多智能体强化学习（MARL）流程，让智能体学习用自然语言进行交流，从而加速有效协议的形成，甚至能够以零样本方式推广到新的队友。其他关于可靠决策的研究 [Lee等人, 2025] 探讨了智能体如何在分布式或不完全信息条件下有效协调。研究表明，智能体之间简单的投票或独立聚合往往优于复杂的迭代反馈循环，因为多轮对话中的误差传播可能使系统不稳定。

不完全信息博弈已得到明确测试，涵盖多议题谈判 [Abdelnabi 等人，2023] 以及扑克和拍卖等博弈 [Hua等人, 2024]。两项研究均表明，尽管朴素的大语言模型常常难以应对庞大的隐藏状态，或偏离博弈论理性，但通过结构化推理，其性能会显著提升。基准测试进一步展现了信息受限的影响。GTBench [Duan等人， 2024] 发现，大语言模型在完全信息确定性博弈（如井字棋）中会彻底失败，但在由机会因素驱动的博弈或随机博弈中仍能保持竞争力。现实风格的模拟也凸显了这些影响。Park 等人 [Park等人, 2023] 表明，25 个生成式智能体能够通过本地记忆和迭代对话，协调完成复杂的社会任务。Piao 等人 [Piao 等人， 2025] 将这一研究扩展到 10,000 个智能体，展示了极化等宏观现象如何从去中心化的交流中涌现。两者共同证明，智能体能够通过智能体间通信，从局部交互中推断全局语境，从而弥补部分可观测性的不足。

4 讨论

本节首先回顾该领域具有代表性的基准测试。随后，第 4.2 节通过分析具体示例，展示博弈论如何助力大语言模型智能体的设计。最后，第 4.3 节指出关键差距，并为该领域提出潜在的研究方向。

4.1 基准测试

基于大语言模型的多智能体系统的基准测试分为两大类，分别侧重于基础能力和专用领域应用。

通用认知基准。 这些基准测试通过将核心能力与领域知识分离，评估协调与推理能力。例如，GAIA要求智能体执行多步骤任务，例如检索历史数据并进行计算，以测试其自主分解问题的能力。

特定领域评估框架。 这类框架将重点转向垂直行业中的功能效用。软件工程代码库或临床诊断模拟器等环境，要求智能体具备深厚的本体知识和专用工具调用能力。在此类评估中，重点在于智能体是否遵守专业协议，以及其执行模拟现实行业需求的协作工作流时的效率。

4.2 案例研究

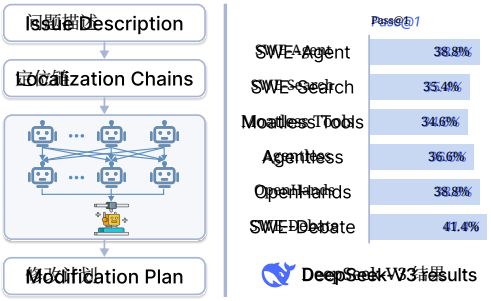


图 4：SWE-Debate 的框架与性能。（左）该工作流采用多智能体辩论机制，迭代优化修改方案。（右）这一竞争式架构达到了最先进水平的 41.4% Pass@1（DeepSeek-V3），显著优于非竞争式基线，并验证了对抗性交互的有效性。

案例一： 多智能体系统中精心设计的竞争机制对自动化软件修复至关重要。

例如，SWE-Debate [Li 等， 2025] 将软件修复建模为非合作博弈，利用多路径故障轨迹作为初始策略，以摆脱单个智能体的局部最优。通过结构化的三轮辩论，智能体被迫进入“辩护与批判”循环，从而揭示潜在的架构权衡。这种对抗性张力使判别器能够综合出可靠的证据，消解歧义，最终达到全局最优均衡。

案例二： 奖励塑形与监管能够使去中心化激励与金融决策中的系统稳定性相一致。

表1：**代表性多智能体系统基准的分类体系**。基准测试按应用领域分组，体现了从通用协调向领域专门化多智能体协作的转变。

类别	领域	代表性基准测试
通用	通用/混合	MultiAgentBench [Zhu 等， 2025]；GAIA [Mialon 等， 2024]；AgentVerse [Chen 等，2024]；Magentic-One [Fourney 等， 2024]
专用	数据科学	MLE-bench [Chan等， 2025]；DSBench [Jing等， 2025]；DABstep [Egg等， 2025]（自主机器学习工程）
	软件工程	SWE-bench [Jimenez 等， 2024]；rSDE-bench [Hu 等， 2024]（仓库级推理）
	金融	FinBen [Xie 等，2024b]；AI-Trader [Fan等， 2025]；FinGAIA [Zeng 等， 2025]（角色-专用决策）
	规划 / 机器人学	TravelPlanner [Xie 等，2024a]；MAP-THOR [Nayak 等， 2024]；PARTNR [Chang 等，2025]（长时程协调）

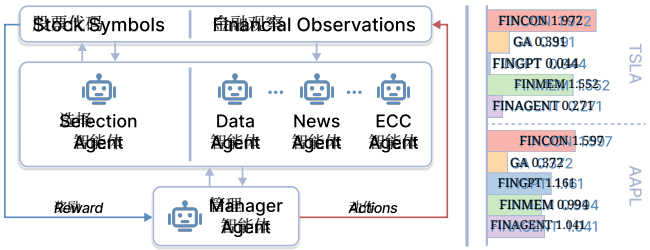


图5：FinCon框架中的奖励塑形。（左）由管理智能体协调的去中心化多智能体金融系统，智能体行为通过基于异构金融观测得出的塑形奖励加以引导。（右）实证比较展示了奖励塑形和监管惩罚如何影响不同资产及骨干模型下的智能体性能。

金融市场本质上是复杂系统，其特点是错综复杂的相互依赖关系；其理论基础深深植根于不对称信息条件下的市场微观结构 [Xu 等， 2025]。在此类环境中，智能体面临私人利润 v_i^{self} 与市场稳定 v^{col}之间的根本冲突，FinCon [Yu 等，2024] 框架表明，调整奖励信号对于确保合乎伦理且稳定的结果至关重要。从博弈论视角来看，通过势函数 $\Phi_i(s)$进行奖励塑形，系统能够加速向合作目标收敛，同时不改变最优策略；而监管机制则纳入惩罚函数 $\Psi_i(s,\ a)$，以遏制资源滥用或伦理违规等不良行为。这些有条理的奖励干预——包括口头反馈或自我惩罚——能够有效引导去中心化大语言模型智能体走向稳定均衡，使个体动机与集体利益保持一致。

4.3 未来方向

分层超级智能体编排。当前的多智能体系统主要依赖预定义协议或静态角色分配。未来研究应转向具备两项核心功能的自主超级智能体：

1.**资源规划**：通过高层元协调，对现有智能体与工具进行理性编排，以解决复杂的长上下文任务。

2. **生成式智能体合成**：根据涌现的环境需求，动态创建量身定制的新型专用智能体。

这些系统必须充当机制设计者，通过协作式综合或有结构的对抗张力来解决架构权衡。

智能体演化。为突破固定策略交互的局限，智能体强化学习是推动策略持续演化的重要范式。研究应将提示选择建模为策略性行动，并识别一种纳入大语言模型有限理性的“推理均衡”。借助竞争性自我博弈和回合级优势估计，智能体可以动态更新参数。这一迭代过程使策略能够在零和与合作型语言博弈中逐步收敛至最优。

理论形式化当前方案依赖启发式实现，但缺乏严格的数学形式化，这限制了对多智能体交互的理论分析。未来工作应建立严谨的框架，以对不完全信息下的决策进行建模，例如借助贝叶斯博弈的视角。一个关键挑战是设计激励相容协议，确保真实信号传递，并使去中心化行动与系统稳定性相一致。

5 结论

我们从博弈论视角对基于大语言模型的多智能体系统进行了全面综述，为这一快速发展的领域提供了统一的理论基础。通过围绕参与者、策略、收益和信息等核心要素梳理当前研究，我们建立了一个系统框架，用以归类多样的智能体行为与交互动态。我们的分析表明，尽管大语言模型擅长高层次推理和策略性沟通，但在复杂的部分可观测环境中，稳健的均衡选择和激励相容性方面仍存在显著差距。最后，我们指出了未来的研究方向，并强调将经典博弈论与大语言模型相结合，对于开发更可靠、自主且具备社会智能的多智能体系统至关重要。

参考文献

[Abdelnabi 等人， 2023] Sahar Abdelnabi、Amr Gomaa、Sarath Sivaprasad 等。Llm-deliberation：通过交互式多智能体谈判游戏评估大语言模型。arXiv 预印本 arXiv:2309.17234，2023。[Agashe 等人， 2025] Saaket Agashe、Yue Fan、Anthony Reyna 和 Xin Eric Wang。LLM-coordination：评估和分析大语言模型的多智能体协调能力。载于 NAACL 会议论文集，第 8038–8057 页，2025。[Ashery 等人， 2025] Ariel Flint Ashery、Luca Maria Aiello 和 Andrea Baronchelli。大语言模型群体中涌现的社会惯例与集体偏见。《科学进展》，11(20):eadu9368，2025。[Bo 等人， 2024] Xiaohe Bo、Zeyu Zhang、Quanyu Dai 等。基于大语言模型的反思式多智能体协作。NeurIPS 会议论文集，37:138595–138631，2024。[Chan 等人， 2025] Jun Shern Chan、Neil Chowdhury、Oliver Jaffe 等。MLE-bench：评估机器学习智能体在机器学习工程任务中的表现。载于 ICLR 会议论文集，2025。[Chang 等人， 2025] Matthew Chang、Gunjan Chhablani、Alexander Clegg 等。PARTNR：用于具身多智能体任务中规划与推理的基准测试。载于 ICLR 会议论文集，2025。[Chen 等人， 2024] Weize Chen、Yusheng Su、Jingwei Zuo 等。AgentVerse：促进多智能体协作并探索涌现行为。载于 ICLR 会议论文集，2024。[Cheng 等人， 2024] Pengyu Cheng、Yong Recognition Dai、Tianhao Hu 等。自我对弈对抗语言游戏增强大语言模型推理能力。NeurIPS 会议论文集，37:126515–126543，2024。[Costarelli 等人， 2024] Anthony Costarelli、Mat Allen、Roman Hauksson 等。GameBench：评估大语言模型智能体的战略推理能力。arXiv 预印本 arXiv:2406.06613，2024。[Duan 等人， 2024] Jinhao Duan、Renming Zhang、James Diffenderfer 等。GTBench：通过博弈论评估揭示大语言模型在战略推理方面的局限。载于 NeurIPS 会议论文集，2024。[Duetting 等人， 2024] Paul Duetting、Vahab Mirrokni、Renato Paes Leme 等。大语言模型的机制设计。载于 ACM Web Conference 会议论文集，第 144–155 页，2024。[Egg 等人， 2025] Alex Egg、Martin Iglesias Goyanes、Friso Kingma 等。DABstep：用于多步推理的数据智能体基准。arXiv 预印本 arXiv:2506.23719，2025。[Epstein， 1999] Joshua M. Epstein。基于智能体的计算模型与生成式社会科学。《复杂性》，4(5):41–60，1999。[(FAIR)† 等人， 2022] Meta 基础人工智能研究外交团队 (FAIR)†、Anton Bakhtin、Noam Brown 等。通过将语言模型与战略推理相结合，在《外交》游戏中达到人类水平的博弈表现。《科学》，378(6624):1067–1074，2022。[Fan 等人， 2025] Tianyu Fan、Yuhao Yang、Yangqin Jiang 等人。《AI-Trader：在实时金融市场中对自主智能体进行基准测试》。arXiv 预印本 arXiv:2512.10971，2025年。[Fourney 等人， 2024] Adam Fourney、Gagan Bansal、Hussein Mozannar 等人。《Magentic-One：用于解决复杂任务的通用多智能体系统》。arXiv 预印本 arXiv:2411.04468，2024年。[Fudenberg and Levine， 1998] Drew Fudenberg 和 David K Levine。《博弈中的学习理论》，第 2 卷。麻省理工学院出版社，1998年。[Hong 等人， 2024] Sirui Hong、Mingchen Zhuge、Jonathan Chen 等人。《MetaGPT：面向多智能体协作框架的元编程》。载于 ICLR 会议论文集，2024年。[Hu 等人， 2024] Yue Hu、Yuzhu Cai、Yaxin Du 等人。《面向软件开发的自进化多智能体协作网络》。arXiv 预印本 arXiv:2410.16946，2024年。[Hua 等人， 2024] Wenyue Hua、Ollie Liu、Lingyao Li 等人。《博弈论大语言模型：谈判博弈的智能体工作流》。arXiv 预印本 arXiv:2411.05990，2024年。[Huang 等人， 2022] Wenlong Huang、Pieter Abbeel、Deepak Pathak 和 Igor Mordatch。《将语言模型用作零样本规划器：为具身智能体提取可执行知识》。载于 ICML 会议论文集，第 9118–9147 页，2022年。[Jimenez 等人， 2024] Carlos E. Jimenez、John Yang、Alexander Wettig 等人。《SWE-bench：语言模型能否解决现实世界中的 GitHub 问题？》载于 ICLR 会议论文集，2024年。[Jing 等人， 2025] Liqiang Jing、Zhehui Huang、Xiaoyang Wang 等人。《DSBench：数据科学智能体距离成为数据科学专家还有多远？》载于 ICLR 会议论文集，2025年。[Kamenica and Gentzkow， 2011] Emir Kamenica 和 Matthew Gentzkow。《贝叶斯劝说》。《美国经济评论》，101(6):2590–2615，2011年。[Lee 等人， 2025] Xian Yeow Lee、Shunichi Akatsuka、Lasitha Vidyaratne 等人。《面向多智能体大语言模型系统的可靠决策》。arXiv 预印本 arXiv:2406.04092，2025 年。[Li 等人， 2023] Guohao Li 、Hasan Abed Al Kader Hammoud、Hani Itani 等人。《CAMEL：用于探索大语言模型社会思维的交互式智能体》。NeurIPS 会议论文集，36:51991–52008，2023年。[Li 等人， 2024] Huao Li、Hossein Nourkhiz Mahjoub、Behdad Chalaki 等人。《基于语言的多智能体强化学习及可由人类理解的通信》。载于 NeurIPS 会议论文集，第 37 卷，第 87908–87933 页，2024年。[Li 等人， 2025] Han Li、Yuling Shi、Shaoxin Lin 等人。《SWE-debate：用于软件问题解决的竞争性多智能体辩论》。arXiv 预印本 arXiv:2507.23348，2025年。[Liang 等人， 2025] Yunhao Liang、Yuan Qu、Jingyuan Yang 等人。《人人贡献！通过序贯公共物品博弈激励多大语言模型系统中的策略性合作》。arXiv 预印本 arXiv:2508.02076，2025年。

[Littman, 1994] Michael L. Littman。《马尔可夫博弈：多智能体强化学习的框架》。载于 ICML 会议论文集，第 157–163 页，1994年。[Liu 等人, 2025] Bo Liu、Leon Guertler、Simon Yu 等人。《SPIRAL：在零和博弈中进行自博弈，通过多智能体多轮强化学习激励推理》。arXiv 预印本 arXiv:2506.24119，2025年。[Ma 等人, 2024] Hao Ma、Tianyi Hu、Zhiqiang Pu 等人。《与另一个自己共同进化：通过序贯协作式多智能体强化学习对大语言模型进行微调》。NeurIPS 会议论文集，37:15497–15525，2024年。[Mas-Colell 等人, 1995] Andreu Mas-Colell、Michael Dennis Whinston、Jerry R Green 等人。《微观经济学理论》，第 1 卷。牛津大学出版社，纽约，1995年。[Mialon 等人, 2024] Gr´egoire Mialon、Cl´ementine Fourrier、Thomas Wolf、Yann LeCun 和 Thomas Scialom。《GAIA：通用人工智能助手的基准测试》。载于 ICLR 会议论文集，2024年。[纳什, 1950] John Nash。《n人博弈的均衡点》。《美国国家科学院院刊》，36(1):48–49，1950年。[Nayak 等人, 2024] Siddharth Nayak、Adelmo Morrison Orozco、Marina Ten Have 等人。《MAP-THOR：对部分可观测环境中的长时域多智能体规划框架进行基准测试》。载于 ICML 2024 多模态基础模型与具身人工智能研讨会论文集，2024年。[Orner 等人, 2025] Maayan Orner、Oleg Maksimov、Akiva Kleinerman 等人。《解释智能体在混合动机博弈中的决策》。载于 AAAI 会议论文集，第 23267–23275 页，2025年。[Park 等人, 2023] Joon Sung Park、Joseph C. O'Brien、Carrie J. Cai 等人。《生成式智能体：人类行为的交互式模拟体》。UIST 会议论文集，第 1–22 页，2023年。[Piao 等人, 2025] Jinghua Piao、Yuwei Yan、Jun Zhang 等人。《AgentSociety：大规模仿真大语言模型驱动的生成式智能体，推进对人类行为与社会的理解》。arXiv 预印本 arXiv:2502.08691，2025年。[Qian 等人, 2024] Chen Qian、Wei Liu、Hongzhang Liu 等人。《ChatDev：用于软件开发的交互式智能体》。载于 ACL 会议论文集，第 15174–15186 页，2024年。[Reid 等人, 2025] Alistair Reid、Simon O'Callaghan、Liam Carroll 和 Tiberio Caetano。《受治理的大语言模型多智能体系统的风险分析技术》。arXiv 预印本 arXiv:2508.05687，2025年。[Sandholm, 1999] Tuomas W. Sandholm。《分布式理性决策》。载于《多智能体系统：分布式人工智能的现代方法》，第 201–258 页。麻省理工学院出版社，1999年。[Shoham and Leyton-Brown, 2008] Yoav Shoham 和 Kevin Leyton-Brown。《多智能体系统：算法、博弈论与逻辑基础》。剑桥大学出版社，2008年。[Silver 等, 2016] David Silver 等。用深度神经网络和树搜索掌握围棋。《自然》，529(7587):484–489，2016。[Wei 等, 2022] Jason Wei、Xuezhi Wang、Dale Schuurmans 等。思维链提示能够激发大语言模型的推理能力。载于《NeurIPS 论文集》，第 35 卷，第 24824–24837 页，2022。[Wu 等, 2024] Qingyun Wu、Gagan Bansal、Jieyu Zhang 等。AutoGen：通过多智能体对话赋能新一代大语言模型应用。载于第一届语言建模会议，2024。[Xie 等，2024a] Jian Xie、Kai Zhang、Jiangjie Chen 等。TravelPlanner：语言智能体在现实世界规划任务中的基准测试。载于《ICML 论文集》，2024。[Xie 等，2024b] Qianqian Xie、Weiguang Han、Zhengyu Chen 等。FinBen：面向大语言模型的综合金融基准测试。《NeurIPS 论文集》，37:95716–95743，2024。[Xu 等, 2025] Yuanjian Xu、Jianing Hao、Guang Zhang 等。Finripple：使大语言模型与金融市场对齐，以感知事件涟漪效应。载于《ACL Findings 论文集》，2025。[Yu 等, 2024] Yangyang Yu、Zhiyuan Yao、Haohang Li 等。FinCon：一种融合概念性语言强化的综合性大语言模型多智能体系统，旨在提升金融决策能力。《NeurIPS 论文集》，37:137010–137045，2024。[Yuan 等, 2025] Huining Yuan、Zelai Xu、Zheyue Tan 等。MARSHAL：通过战略型大语言模型自博弈激励多智能体推理。arXiv 预印本，arXiv:2510.15414，2025。[Zeng 等, 2025] Lingfeng Zeng、Fangqi Lou、Zixuan Wang 等。FinGAIA：面向现实金融领域人工智能智能体的中文基准测试。arXiv 预印本，arXiv:2507.17186，2025。[Zhang 等，2024a] Ceyao Zhang、Kaijie Yang、Siyi Hu 等。ProAgent：利用大语言模型构建主动协作智能体。载于《AAAI 论文集》，第 17591–17599 页，2024。[Zhang 等，2024b] Ruize Zhang、Zelai Xu、Chengdong Ma 等。强化学习中自博弈方法综述。arXiv 预印本，arXiv:2408.01072，2024。[Zhang 等，2024c] Yusen Zhang、Ruoxi Sun、Yanfei Chen 等。智能体链：大语言模型在长上下文任务中的协作。《NeurIPS 论文集》，37:132208–132237，2024。[Zhu 等, 2025] Kunlun Zhu、Hongyi Du、Zhaochen Hong 等。MultiAgentBench：评估大语言模型智能体的协作与竞争。载于《ACL 论文集》，第 8580–8622 页，2025。[Zhu, 2025] Quanyan Zhu。大语言模型纳什博弈中的推理与行为均衡：从思维模式到行动。arXiv 预印本，arXiv:2507.08208，2025。