中国科学 : 信息科学 2026 年 第 56 卷 第 2 期: 485–486
SCIENTIA SINICA Informationis
《中国科学》杂志社
SCIENCE CHINA PRESS
精选文章导读

基于大语言模型的智能体: 发展与未来展望

奚志恒[1], 陈文翔[1], 郭昕[1], 何为[1], 丁怡文[1], 洪博杨[1], 张明[1], 王浚哲[1], 金森杰[1], 周恩宇[1], 郑锐[1], 樊晓然[1], 王枭[1], 熊立茂[1], 周钰皓[1], 王威然[2], 江常皓[1], 邹易澄[1], 刘向阳[1], 印张悦[1], 窦士涵[1], 翁荣祥[4], 秦文娟[2], 郑咏滟[2], 邱锡鹏[1], 黄萱菁[1], 张奇[1*], 桂韬[3*]

1. 复旦大学计算与智能创新学院, 上海 200441
2. 复旦大学外国语言文学学院, 上海 200433
3. 复旦大学现代语言学研究院, 上海 200433
4. 美团, 北京 100102
* 通信作者. E-mail: qz@fudan.edu.cn, tgui@fudan.edu.cn

长期以来, 实现通用类人智能始终是人类不懈追求的目标 [1]. 而智能体 (agent), 即能够感知环境、作出决策并采取行动的人工实体, 被视为实现该目标的重要载体. 尽管目前已有大量智能体相关的研究, 但这些研究大多集中于算法或训练方法的改进, 旨在增强智能体在特定任务上的能力或性能. 事实上, 研究社区目前正缺乏一个通用而强大的模型, 作为打造能够适应多种场景的智能体的基础. 近年来, 大语言模型 (LLMs) 凭借其强大的通用能力, 为实现通用人工智能 (AGI) 带来了新的希望. 许多研究者已将大语言模型作为构建智能体的核心基础, 并取得了显著进展 [2~4].

*SCIENCE CHINA Information Sciences*在第 68 卷第 2 期出版了 “The rise and potential of large language model based agents: a survey”, 系统回顾并分析了精心筛选的 574 篇相关文献, 旨在全面梳理基于大语言模型的智能体的发展、通用框架、应用场景与智能体社会的构建, 同时指出该领域未来可能的发展方向与尚未解决的关键问题.

首先, 从历史演进的角度, 梳理了智能体研究从基于规则的符号系统, 到数据驱动的机器学习方法, 再到当前以大语言模型为核心的演进路径, 从自主性 (autonomy)、反应性 (reactivity)、主动性 (proactiveness) 和社会性 (social ability) 4 个方面阐明了大语言模型作为智能体的核心所带来的根本性变革. 接着, 文章提出并详细阐述了一个基于大语言模型的智能体通用框架, 包含大脑 (brain)、感知 (perception) 和行动 (action) 三大核心组成部分. 其中, 感知模块负责将多模态环境信息 (文本、视觉、听觉等) 转化为大语言模型可理解的表示; 大脑则负责基于感知信息进行推理 (reasoning) 和规划 (planning), 同时维护外部知识 (knowledge) 与内部记忆 (memory), 具备良好的迁移 (transferability) 和泛化 (generalization) 能力; 行动模块则将决策转化为具体的语言/动作输出、工具使用 (tool using) 或是具身行为 (embodied action). 这一框架为理解和设计各类大语言模型智能体提供了统一的概念基础.

在应用场景方面, 文章分别从单智能体 (single agent)、多智能体 (multiple agents) 和人机交互 (human-agent interaction) 3 个角度系统性地展示了基于大语言模型的智能体在众多领域的落地潜力与初步成果.

- 单智能体: 单智能体已经在广泛的任务和场景下表现出了强大的通用能力, 文章将其概括为面向任务 (task-oriented)、面向研究创新 (innovation-oriented) 与面向生命模拟 (lifecycle-oriented) 三大应用场景. 对于

英文原文: Xi Z H, Chen W X, Guo X, et al. The rise and potential of large language model based agents: a survey. Sci China Inf Sci, 2025, 68: 121101, doi: 10.1007/s11432-024-4222-0

面向任务的应用, 文章指出智能体需要合理地将任务划分为多个子任务并逐步完成, 例如在网页场景下能够理解需求、自动生成代码和调试程序; 而生活场景下可以提供个性化的日程管理、信息检索、旅行规划、健康咨询等全方位服务. 对于**面向研究创新**的应用, 文章主要阐述了科学研究、技术开发和创新性任务解决等方面, 例如在科学研究领域, 智能体可以辅助文献调研、实验设计甚至代码编写与数据分析. 对于**面向生命模拟**的应用, 文章强调这类智能体不仅需要理解和生成自然语言, 还需要具备一定的常识和社会认知能力, 并展示出智能体在模拟生存环境中惊人的持续探索和开发新技能的能力.

• 多智能体: 在大模型智能体的应用中, 多智能体的协作范式通常包括合作互动 (cooperative interation) 和对抗互动 (adversarial interaction) 这两类主要模式. 而对于**合作互动**模式, 具体又可根据任务执行是否有序以及是否涉及角色分工 (role-playing) 将其划分为无序合作 (disordered cooperation) 与有序合作 (ordered cooperation) 两类. 在这种模式下, 智能体通过扮演各自不同的角色进行分工交流, 共享资源和信息, 从而更高效地完成任务. 对于**对抗互动**模式, 即在智能体之间设置对抗性任务和竞争环境, 从而实现整体性能的提升, 其重要的思想在于引入辩论 (debate) 机制. 在这种机制下, 每个智能体承担不同的角色, 并提出各自的观点或解决方案. 通过设定明确的辩论规则和评价标准, 智能体之间展开有序的辩论. 这种模式能够帮助智能体发现自身的缺陷和不足, 从而进行自我改进.

• 人机交互: 智能体的最终目的是要服务于人类, 因此, 人机交互范式通过在系统中引入人类的参与实现人机智能互动. 进一步地, 人机交互可以分为人类主导和人机平等两类协作范式. 在**人类主导**的范式中, 人类对智能体提供适当的指导与反馈, 从而主导智能体的行为和决策. 根据人类的反馈类型, 又可以进一步分为定量 (quantitative) 与定性 (qualitative) 反馈. 这种模式强调人类的主导地位, 智能体在执行任务时依赖于人类的指示与纠正. 而**人机平等**的范式, 强调智能体与人类作为平等的合作伙伴, 共同参与任务的规划和执行. 这种模式强调智能体的适应性, 通过与人类的协同工作高效完成任务, 同时通过自主学习提高自身能力. 随着智能体的基础环境感知与推理能力的提升, 双方的互动方式和理解将不断优化, 实现真正的无缝协同.

特别地, 文章探讨了智能体社会这一前沿方向, 从智能体的社会行为与人格、模拟社会的运行环境, 以及智能体的社会模拟 3 个层次展开. 首先, 从外部维度观察智能体的社会行为 (social behavior), 包括智能体如何独立行动, 以及如何与其所处环境互动; 另一方面探究智能体表现出的人格 (personality), 例如认知、情感、性格等, 这些都会影响到智能体的行为. 其次, 文章分别讨论了基于文本的模拟环境 (text-based)、虚拟的沙盒环境 (virtual sandbox) 以及真实的物理环境 (physical). 最后, 当大量具有自主性的大模型智能体共存于一个共享环境时, 便形成了初步的智能体社会. 目前, 针对模拟社会的研究主要从两个方面进行, 一是探索智能体的群体能力边界, 二是利用模拟社会来进行社会学的研究. 研究者们通过构建多智能体社会, 观察到了交流、合作、社会关系形成乃至文化现象涌现等复杂行为. 文章分别从开放性 (open)、持久性 (persisitent)、情境性 (situated) 和组织性 (organized) 四个特性来分析模拟社会的运行方式, 并总结了智能体社会的现象与启示, 同时指出潜在的问题与风险.

进一步地, 文章对基于大模型的智能体进行了深刻的讨论, 包括大语言模型和智能体研究领域的互相促进关系、大模型智能体的开发框架、评估大模型智能体能力的维度等. 最后, 文章指出大语言智能体在安全性、可靠性等方面仍面临一系列关键挑战, 同时面向未来给出了可能的发展方向. 基于大语言模型的智能体研究正在开启人工智能发展的新篇章, 未来更加通用、可靠、高效的智能体将推动我们向着通用人工智能的宏伟目标稳步迈进.

参考文献

1 Russell S. Artificial Intelligence: A Modern Approach. Upper Saddle River: Pearson Education, Inc., 2020

2 Park J S, O'Brien J C, Cai C J, et al. Generative agents: interactive simulacra of human behavior. 2023. ArXiv:2304.03442

3 Li G, Hammoud H A A K, Itani H, et al. CAMEL: communicative agents for "mind" exploration of large scale language model society. 2023. ArXiv:2303.17760

4 Wang G, Xie Y, Jiang Y, et al. Voyager: an open-ended embodied agent with large language models. 2023. ArXiv:2305.16291