

本文引用格式: 王航,黄小平,甘国操,等.基于大型语言模型的异构多智能体系统[J].自动化与信息工程,2025,46(4):22-27.

WANG Hang, HUANG Xiaoping, GAN Guocao, et al. Heterogeneous multi-agent system based on large language models[J]. Automation & Information Engineering, 2025,46(4):22-27.

基于大型语言模型的异构多智能体系统*

王航 黄小平 甘国操 冉瑛

(中国民航信息网络股份有限公司重庆分公司, 重庆 401122)

摘要: 针对企业在数字化转型过程中, 采用单一智能体框架实现多业务场景智能化时面临的困境, 提出一种基于大型语言模型的异构多智能体系统。该系统基于企业业务领域划分智能体, 并构建一种异构多智能体系统分层架构, 使智能体系统以分层的方式协同工作。其中, 主控智能体负责业务领域识别、服务注册、服务分发和结果汇集等工作; 各领域智能体以微服务架构独立运行, 提升了异构多智能体系统的效率和灵活性。通过实验验证了该系统在众多且差异化较大的场景下, 不仅能够完成特定任务, 还能够自主选择执行效果最好的智能体。

关键词: 异构多智能体系统; 大型语言模型; 主控智能体; 领域智能体; 微服务架构

中图分类号: TP18

文献标志码: A

文章编号: 1674-2605(2025)04-0003-06

DOI: 10.12475/aie.20250403

开放获取

Heterogeneous Multi-agent System Based on Large Language Models

WANG Hang HUANG Xiaoping GAN Guocao RAN Ying

(China Travelsky Information Technology Co., Ltd., Chongqing Branch, Chongqing 401122, China)

Abstract: Aiming at the challenges faced by enterprises during digital transformation when adopting a single-agent framework to achieve intelligent solutions across multiple business scenarios, this paper proposes a heterogeneous multi-agent system based on large language models. The system categorizes agents based on enterprise business domains and constructs a hierarchical architecture for heterogeneous multi-agent systems, enabling agents system to collaborate in a layered manner. Within this architecture, a master agent is responsible for tasks such as business domain recognition, service registration, service distribution, and result aggregation. Domain-specific agents operate independently based on a microservices architecture, enhancing the efficiency and flexibility of the heterogeneous multi-agent system. Experimental results demonstrate that the system not only accomplishes specific tasks across numerous and highly diverse scenarios but also autonomously selects the agent with the best execution performance.

Keywords: heterogeneous multi-agent system; large language models; master agent; domain-specific agents; microservices architecture

0 引言

随着人工智能技术的快速发展, 大型语言模型 (large language model, LLM)^[1], 如 ChatGPT^[2]、Deep-Seek^[3]等, 以其卓越的自然语言理解、处理和生成能力, 在多个领域展现出广阔的应用前景。依托 LLM 构建的智能体^[4]能够基于用户意图自主执行决策和行动, 并可以整合多个具有不同能力的智能体, 形成多智能体系统^[5]。多智能体系统中的每个智能体专注于特定

的任务, 通过彼此之间的交互, 能更高效、智能地共同完成复杂任务。

目前, 多智能体系统已相继出现多种开源框架, 如 MetaGPT^[6]、AutoGen^[7]、LangGraph^[8]、CrewAI^[9]等, 在智能体协作、任务分解、信息共享等方面凸显出优势。其中, MetaGPT 通过模拟和优化智能体的行为, 能够处理复杂的软件开发等任务; AutoGen 中每个智能体专注于不同的编程任务, 广泛应用于数学问

22 * 基金项目: 佛山市顺德区科技创新项目(2130218002519); 广东省科学院青年人才专项(2024GDASQNR-0327)。

题求解、代码生成、决策制定等领域；LangGraph 采用图形化表示法建模智能体之间的协作关系，使用户可直观地查看并调整智能体之间的协作关系，实现高效的协作和任务分配；CrewAI 是一个基于角色和流程的多智能体框架，支持复杂的任务分解和流程管理，能够确保各智能体按照预定流程有序地执行任务。然而，上述开源框架大多聚焦于单一框架的智能体协作，适用场景有限。同时，企业在选择智能体框架时可能面临以下问题：1) 企业与第三方服务（如钉钉、豆包、腾讯、百度等）进行 AI 集成时，技术栈通常是异构的，单一智能体框架难以满足需求；2) 单一智能体框架无法满足不同业务场景（如办公、研发、客服等）的需求；3) 当前智能体的组织形式与企业的组织架构无法适配。

为此，本文提出一种基于 LLM 的异构多智能体系统，为企业不同业务场景的全面智能化提供一种有效兼容不同智能体框架的解决方案，从而推动企业的数字化转型^[10]。

1 相关内容

1.1 多智能体控制机制

多智能体控制机制借鉴了多机器人系统的组织架构^[11]，主要包括集中式控制、分布式控制、混合控制等 3 种机制。其中，集中式控制的决策权和控制权集中在中心智能体，其本质是建立一种垂直层级的权力结构，即下级智能体执行上级智能体的决策和指示；分布式控制的每个智能体都具有独立的决策权和控制权，通过通信网络实现智能体之间的信息共享、协调工作；混合控制是集中式控制和分布式控制的结合体，既保留了中心智能体的决策权，又充分发挥了各智能体的自主性和灵活性。

1.2 多智能体协作机制

多智能体协作机制主要描述和分析智能体之间的交互与协作关系，可分为基于角色、基于协商和基于学习 3 种模型。其中，基于角色的模型将人类社会的角色概念引入多智能体系统，以描述智能体之间的交互关系；基于协商的模型通过智能体之间的信息交

换来达成共识或解决冲突；基于学习的模型强调智能体通过学习（如强化学习、深度学习等）和适应来改进自身行为。

1.3 智能体架构模式

智能体是指能够感知环境并自主采取行动以实现特定目标的代理体，其表现形式可以是软件、硬件或系统。常见的软件构架模式包括单体应用架构、微服务架构^[12]、面向服务架构（service-oriented architecture, SOA）等。其中，单体应用架构将应用程序作为一个不可分割的整体进行开发设计，所有的功能模块、数据库、第三方服务等都被集成在一个代码库中，构成一个庞大的应用程序；微服务架构将单体应用划分为一组服务，每个服务都运行在独立的进程中，通过轻量级通信机制进行服务间协作；SOA 作为一种组件模型，将应用程序的不同功能单元进行拆分，并通过这些服务间预定义的接口和协议进行联系。

1.4 智能体认知框架

智能体认知框架是指智能体在执行任务时依赖的思维模式和决策机制，主要包括思维链（chain of thought, COT）^[13]、ReAct^[14]、Function Call^[15]、Plan-and-Executor^[16]等。其中，COT 通过将逻辑复杂的问题进行拆解，形成一系列具有逻辑关系的思考步骤，从而构建完整的思考过程；ReAct 以交替的方式结合智能体的推理能力和行动能力，借助动态决策和上下文感知，实现更加智能和高效的交互；Function Call 能够调用特定的函数或接口来增强智能体处理任务的能力；Plan-and-Executor 强调智能体的任务规划与执行分离，并允许智能体在执行过程中根据反馈动态调整任务规划。

2 系统设计

2.1 设计思路

基于 LLM 的异构多智能体系统设计思路如下：

1) 根据企业业务领域的描述、角色定义和协作分工，采用基于角色的模型来设计智能体，以适应不同的业务领域；

2) 不同业务领域能够自适应地选择智能体框架, 而同一业务领域的多个智能体, 其框架选择策略需综合考虑任务耗时、成功率等因素;

3) 由于基于角色的智能体之间存在能力差异, 因此采用混合控制机制, 并根据业务领域特点选择合适的框架;

4) 考虑到需要集成第三方服务, 以及各业务领域之间的有效协同, 选择微服务架构来组织智能体。

2.2 系统架构设计

基于 LLM 的异构多智能体系统采用分层架构, 主要包括用户层、主控层、领域层、基础设施层, 如图 1 所示。

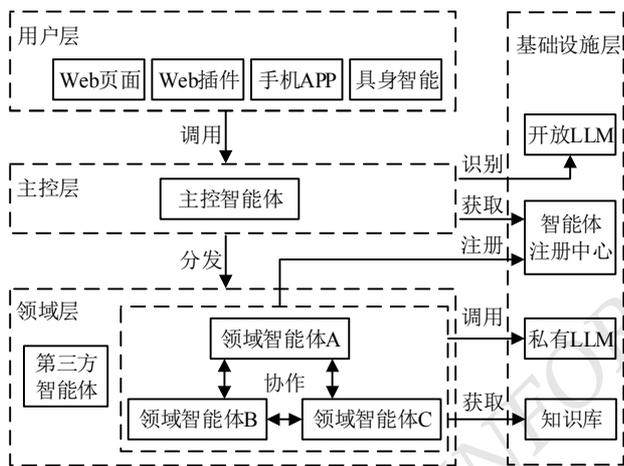


图 1 基于 LLM 的异构多智能体系统架构

用户层主要为用户提供多种操作入口, 主要包括 Web 页面、Web 插件、手机 APP、具身智能^[17-18]等接入方式。

主控层作为系统的核心部分, 主要负责敏感词识别与过滤、同义词转换、领域识别、服务注册、服务分发和结果汇集等工作。

领域层包含多个智能体微服务架构, 负责接收主控层分发的任务, 并与其他智能体协作完成复杂任务。同时, 该层支持不同框架构建的领域智能体, 各领域智能体之间通过信息共享与协作来完成任务。

基础设施层提供私有 LLM 运行环境、企业知识库, 以满足垂直领域的知识增强和微调需求。

2.3 服务注册

基于 LLM 的异构多智能体系统通过智能体注册中心实现不同领域智能体的服务注册、监控和更新, 其设计图如图 2 所示。

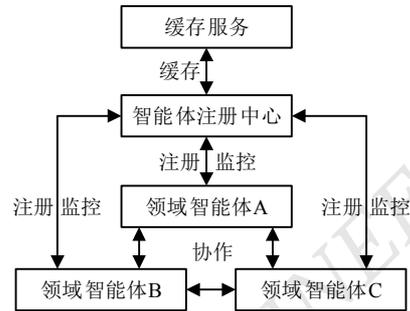


图 2 服务注册设计图

智能体注册中心提供微服务注册功能, 以及统一的 Restful 风格服务注册接口。其接口内容包括归属领域、领域职责描述、Restful 风格的服务调用地址和健康检查地址等信息。通过缓存服务保存智能体信息, 检测智能体服务状态、更新智能体服务目录。各领域智能体在发布时, 调用服务注册接口并自动注册到智能体注册中心, 以便在后续的服务分发过程中, 主控智能体能够获取领域智能体的服务调用地址。如果领域智能体服务下线, 智能体注册中心会基于心跳检测结果, 自动更新服务目录。通过智能体注册中心, 系统能够有效兼容异构多智能体, 解决了企业不同业务场景的协同开发问题。

2.4 服务分发

基于 LLM 的异构多智能体系统的服务分发机制是系统高效运转的关键, 服务分发设计图如图 3 所示。

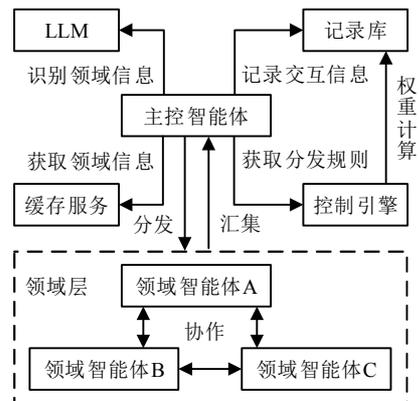


图 3 服务分发设计图

系统收到用户请求后，主控智能体从智能体注册中心的缓存服务中获取智能体的服务目录；基于 LLM 识别用户请求涉及的领域，并进行任务规划；根据控制引擎提供的分发规则，将规划任务分发给相应的领域智能体；当各领域智能体完成任务后，主控智能体汇集所有结果，借助 LLM 汇总并返回。

当同一业务领域注册了多个不同框架的智能体时，控制引擎会定时计算这些智能体的任务耗时、成功率等指标数据，并将这些指标数据保存到记录库。通过对这些指标数据进行加权计算，控制引擎获得相应领域智能体的分发权重，并按照最高权重生成分发规则，实现同一业务领域不同框架智能体的最优选择。

3 实验验证

本文通过实验场景验证基于 LLM 的异构多智能体系统的性能。

3.1 实验环境

实验环境配置如表 1 所示。

表 1 实验环境配置

类型	供应商	模型/配置
开放大模型平台 1	智普 AI 开放平台	GLM-4-AirX
开放大模型平台 2	千帆大模型平台	ERNIE-4.0-8K-0613
私有大模型平台	华为 NPU 服务器	GLM-4/192 C、756 GB、NPU 显卡为 32 GB×8
kubernetes 集群	KVM	3 台 8 C、16 GB、200 GB 磁盘 5 台 64 C、128 GB、500 GB 磁盘

3.2 实验设计与实施

为验证基于 LLM 的异构多智能体系统在框架异构、场景众多且差异化大、跨领域协同等方面的性能，设计了单领域多框架、多领域单框架、多领域多框架 3 个实验场景。采用任务的平均耗时、成功率作为评价指标。

3.2.1 单领域多框架实验场景

为验证不同框架的智能体在单个领域的实施效果，采用 4 种框架构建人力领域智能体，如表 2 所示。

表 2 单领域多框架实验场景

框架	领域	设计（角色/流程/任务）
MetaGPT	人力	总监、经理、人力专员、招聘主管、培训主管、考核主管、薪资主管等角色
AutoGen	人力	组织管理、员工管理、薪资管理、招聘管理、培训管理、绩效管理等流程
CrewAI	人力	组织查询、员工查询、培训管理、绩效计算、薪资计算等任务
LangGraph	人力	总监、经理、人力专员、招聘主管、培训主管、考核主管、薪资主管等角色

3.2.2 多领域单框架实验场景

为了验证单框架的智能体在多个领域的实施效果，分别采用 4 种框架构建人力、行政、财务、党群领域智能体，如表 3 所示。

表 3 多领域单框架实验场景

框架	领域	设计
MetaGPT	人力	总监、经理、人力专员、招聘主管、培训主管、考核主管、薪资主管等角色
	行政	总监、经理、行政主管、行政专员、行政助理等角色
	财务	总监、经理、会计主管、出纳主管、会计、出纳等角色
	党群	部长、组织委员、宣传委员、工会主席、团委书记等角色
AutoGen	人力	组织管理、员工管理、薪资管理、招聘管理、培训管理、绩效管理等流程
	行政	值班管理、档案管理、制度管理、会议管理、资产管理等流程
	财务	报销管理、成本管理、差旅管理、资金管理、费用管理、预算管理等流程
	党群	党员管理、干部管理、工会管理、团委管理、新媒体管理等流程
CrewAI	人力	组织查询、员工查询、培训管理、绩效计算、薪资计算等任务
	行政	值班计划生成、值班数据查询、档案信息查询、制度查询、资产查询等任务
	财务	财务报销、成本核算、会计核算、财务预算等任务
	党群	党员查询、团员查询、工会活动生成和查询等任务
LangGraph	人力	总监、经理、人力专员、招聘主管、培训主管、考核主管、薪资主管等角色
	行政	总监、经理、行政主管、行政专员、行政助理等角色
	财务	总监、经理、会计主管、出纳主管、会计、出纳等角色
	党群	部长、组织委员、宣传委员、工会主席、团委书记等角色

3.2.3 多领域多框架实验场景

为了验证不同框架的智能体在多个领域的实施效果，分别采用4种框架构建人力、行政、财务、党群领域智能体，如表4所示。

表4 多领域多框架实验场景

框架	领域	设计
MetaGPT	人力	总监、经理、人力专员、招聘主管、培训主管、考核主管、薪资主管等角色
AutoGen	行政	值班管理、档案管理、制度管理、会议管理、资产管理等流程
CrewAI	财务	财务报销、成本核算、会计核算、财务预算等任务
LangGraph	党群	部长、组织委员、宣传委员、工会主席、团委书记等角色

3.2.4 实验结果

基于上述3个实验场景的实验结果分别如表5、6、7所示。

表5 单领域多框架实验结果

框架	领域	任务总数/个	平均耗时/s	成功率/%
MetaGPT	人力	100	103	87
AutoGen	人力	100	49	73
CrewAI	人力	100	35	78
LangGraph	人力	100	60	74

表6 多领域单框架实验结果

框架	领域	任务总数/个	平均耗时/s	成功率/%
MetaGPT	人力	100	112	85
	行政	100	107	76
	财务	100	121	72
	党群	100	97	81
AutoGen	人力	100	54	72
	行政	100	51	75
	财务	100	67	66
	党群	100	49	76
CrewAI	人力	100	38	76
	行政	100	49	72
	财务	100	51	73
	党群	100	46	69
LangGraph	人力	100	67	67
	行政	100	62	73
	财务	100	73	65
	党群	100	54	79

由表5、6可知：MetaGPT的任务成功率最高，但平均耗时最长，适用于对实时性要求不高但确定性要求较高的场景；CrewAI的平均耗时最短，任务成功率较高，适用于对实时性和确定性都有一定要求的场景。因此，各领域应该根据其业务特性来选择合适的智能体框架构建智能体。

表7 多领域多框架实验结果

框架	领域	任务总数/个	平均耗时/s	成功率/%
MetaGPT	人力	100	115	85
AutoGen	行政	100	49	77
CrewAI	财务	100	52	74
LangGraph	党群	100	45	81

由表7可知，基于指标数据计算的最高权重生成的分发规则，能够实现不同智能体框架的最优选择。

3.3 实验结果分析

上述实验结果表明：基于不同框架的异构多智能体能够正确完成大多数任务；异构框架的效果比单一框架更优，任务成功率均在74%以上，证明了本文提出的系统不仅能够完成任务，还能够自主选择执行效果最好的智能体。

4 结论

本文提出一种基于LLM的异构多智能体系统，在场景众多且差异化较大的情况下，实现了异构多智能体的协同工作。通过实验验证了该系统在协作效率、成功率、跨领域协同等方面的性能。然而，目前异构多智能体系统在模型选择、角色拆分、任务规划、流程设计等方面仍存在不足，导致任务成功率还不高，因此，需要对异构多智能体系统进行不断优化和调整。

©The author(s) 2024. This is an open access article under the CC BY-NC-ND 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

参考文献

- [1] 舒文韬,李睿潇,孙天祥,等.大型语言模型:原理、实现与发展[J].计算机研究与发展,2024,61(2):351-361.
- [2] 朱光辉,王喜文.ChatGPT的运行模式、关键技术及未来图景[J].新疆师范大学学报(哲学社会科学版),2023,44(4):113-

- 114.
- [3] BI X, CHEN D, CHEN G, et al. DeepSeek LLM: Scaling open-source language models with longtermism[J]. arXiv preprint arXiv:2401.02954,2024.
- [4] XI Z, CHEN W, GUO X, et al. The rise and potential of large language model based agents: A survey[J]. Science China Information Sciences, 2025, 68(2): 121101.
- [5] 李杨,徐峰,谢光强,等.多智能体技术发展及其应用综述[J].计算机工程与应用,2018,54(9):13-21.
- [6] HONG S, ZHUGE M, CHEN J, et al. MetaGPT: Meta programming for a multi-agent collaborative framework[J]. arXiv preprint arXiv:2308.00352,2023,3(4): 6.
- [7] WU Q, BANSAL G, ZHANG J, et al. AutoGen: Enabling Next-Gen LLM applications via multi-agent conversation[J]. arXiv preprint arXiv:2308.08155,2023.
- [8] DUAN Z, WANG J. Exploration of LLM multi-agent application implementation based on LangGraph+ CrewAI[J]. arXiv preprint arXiv:2411.18241,2024.
- [9] Berti A, Maatallah M, Jessen U, et al. Re-thinking process mining in the AI-based agents era[J]. arXiv preprint arXiv: 2408.07720,2024.
- [10] 戚聿东,肖旭.数字经济时代的企业管理变革[J].管理世界, 2020,36(6):135-152;250.
- [11] 吴军,徐昕,连传强,等.协作多机器人系统研究进展综述[J].智能系统学报,2011,6(1):13-27.
- [12] 冯志勇,徐砚伟,薛霄,等.微服务技术发展的现状与展望[J].计算机研究与发展,2020,57(5):1103-1122.
- [13] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. arXiv preprint arXiv:2201.11903,2022.
- [14] YAO S, ZHAO J, YU D, et al. ReAct: Synergizing reasoning and acting in language models[J]. arXiv preprint arXiv:2210.03629,2022.
- [15] 吴迪.基于 Cougaar 的智能 Agent 研究与实现[D].哈尔滨:哈尔滨工业大学,2006.
- [16] ZHAO X, JING M, WU Y. Agent can say no: Robot task planning by natural language feedback between planner and executor[C]//International Conference on Intelligent Computing. Springer, Singapore, 2024:142-153.
- [17] 白辰甲,许华哲,李学龙.大模型驱动的具身智能:发展与挑战[J].中国科学:信息科学,2024,54(9):2035-2082.
- [18] 何文卿,郑立京,李逸,等.基于文献计量的人形机器人具身智能领域现状与发展趋势[J].机电工程技术,2025,54(13): 16-23.

作者简介:

王航,男,1982年生,硕士研究生,工程师,主要研究方向:软件工程、DEVOPS、人工智能。E-mail: hangwang@travelsky.com.cn

黄小平,男,1990年生,本科,工程师,主要研究方向:软件工程、DEVOPS、人工智能。E-mail: hxp@travelsky.com.cn

甘国操,男,1983年生,本科,高级工程师,主要研究方向:信息安全、数字化。E-mail: gcgan@travelsky.com.cn

冉瑛,男,1987年生,本科,高级工程师,主要研究方向:人工智能、大数据。E-mail: ranying@travelsky.com.cn

《自动化与信息工程》期刊公众号、青年编委申请表及审稿专家信息表



期刊公众号



青年编委申请表



审稿专家信息表