

基于大模型的 AI 智能体的构建与创新场景应用探索

刘晓宇

(北京联合大学, 北京)

摘要: 本文聚焦于基于大模型的 AI 智能体的构建及其创新场景应用研究。通过梳理国内外研究现状, 呈现 OpenAI、Google、DeepMind 等国外机构与清华、北大等国内高校及企业在该领域的前沿探索成果。详细剖析 AI 智能体基础理论, 包括其定义、特性、分类, 以及大模型技术两者融合机制。深入探讨智能体构建技术, 涵盖构建框架、关键组件、工具集成及多智能体系统构建。结合医疗、金融、教育、制造业等行业实际案例, 深度解析智能体应用模式与价值, 并对智能家居、娱乐等新兴场景进行展望。最后, 识别研究面临的挑战, 提出针对性解决方案与未来发展方向, 为推动基于大模型的 AI 智能体技术进步与广泛应用提供理论与实践参考。

关键词: 大模型; AI 智能体; 构建技术; 创新场景; 行业案例

中图分类号: N8

文献标识码: A

一、引言

1.1 研究背景

人工智能自 20 世纪 50 年代诞生以来, 经历了多个发展阶段, 从早期基于规则的系统, 到机器学习、深度学习的兴起, AI 技术不断演进, 逐步渗透到人们生活与社会生产的各个领域。随着数据量的爆发式增长和计算能力的提升, 机器学习尤其是深度学习技术得到迅猛发展, 通过大量数据的训练, 模型能够自动学习数据中的模式和特征, 在图像识别、语音识别、自然语言处理等领域取得了显著成果。

大模型的兴起是 AI 发展历程中的一个重要里程碑。大模型通常基于 Transformer 架构, 利用海量的数据进行预训练, 从而具备强大的语言理解与生成能力, 以及广泛的知识储备。例如 OpenAI 的 GPT 系列, 在自然语言处理任务中展现出令人瞩目的性能, 能够生成连贯、逻辑清晰的文本, 在对话交互、内容创作、智能客服等场景中得到广泛应用; 谷歌的 BERT 模型则在语义理解方面取得突破, 推动了自然语言处理任务的精度提升。大模型为 AI 智能体的构建带来了新的契机。基于大模型构建的 AI 智能体, 能够借助大模型强大的语言处理和推理能力, 更好地理解复杂任务, 自主制定决策和行动方案, 实现从感知、理解到决策、执行的全流程智能化。这使得 AI 智能体在面对多样化、动态变化的环境时, 能够展现出更强大的适应性和智能水平。

1.2 国内外研究现状

在大模型研究方面, 国外一直处于领先地位。OpenAI 的 GPT 系列模型无疑是最为瞩目的, GPT-4 展现出在多模态理解、复杂推理、代码生成等多方面的卓越能力, 被广泛应用于智能写作、智能编程、智能客服等多个领域。谷歌的 BERT 模型在自然语言处理的基础研究中贡献巨大, 后续又推出了基于 Transformer 架构的一系列模型, 不断探索大模型在不同场景下的优化与应用, 如在搜索引擎优化、知识图谱构建等方面取得进展。Meta 的 LLaMA 模型以其开源特性, 吸引了全球开发者参与模型的改进与应用开发, 推动了大模型技术的普及与创新, 基于 LLaMA 的各种微调模型在不同语言和任务场景下展现出独特优势。

国内的大模型研究也取得了显著成果。百度的文心一言在自然语言处理和知识图谱融合方面具有特色, 通过对大量中文文本数据的学习, 在中文语言理解与生成任务上表现出色, 尤其在智能搜索、内容创作辅助等场景中得到广泛应用。科大讯飞的星火认知大模型在语音

识别、自然语言生成与交互方面优势明显，结合其在智能语音领域的深厚积累，为智能教育、智能医疗等行业提供了智能化解决方案。此外，华为的盘古大模型、阿里的通义千问等也在各自擅长的领域不断探索创新，如盘古大模型在工业领域的应用，通义千问在电商场景的应用等。

在 AI 智能体构建与应用方面，国外同样有诸多先进的研究与实践。OpenAI 正在探索基于大模型的智能体在复杂任务中的应用，如智能体在模拟环境中的自主学习与决策，以实现更高级别的人工智能。DeepMind 的智能体研究则侧重于强化学习与大模型的结合，通过不断的试错学习，智能体在游戏、机器人控制等领域取得了良好的成果，如 AlphaGo 在围棋领域战胜人类顶尖棋手，展示了智能体在策略学习方面的强大能力。

国内在 AI 智能体领域也在积极跟进。一些研究机构和企业致力于将大模型与智能体技术相结合，开发具有自主决策能力的智能体系统。例如，在智能制造领域，智能体被用于生产线的优化调度，根据实时生产数据和环境变化，自主调整生产策略，提高生产效率和质量；在智能物流领域，智能体能够根据订单信息、车辆位置、交通状况等多源数据，合理规划物流配送路线，实现高效的物流运输。

1.3 研究目的

本研究旨在深入探索基于大模型的 AI 智能体的构建技术，并全面挖掘其在创新场景中的应用潜力。通过对大模型与 AI 智能体相关技术文献资料的研究，探索为构建性能优越、适应性强的 AI 智能体系统提供思路，并将其应用于具有挑战性和创新性的场景中，推动人工智能技术在实际应用中的进一步发展。

二、基于大模型的 AI 智能体基础理论

2.1 AI 智能体概述

2.1.1 AI 智能体的定义

AI 智能体是一种能够感知环境、自主决策并执行行动以实现特定目标的智能实体。它集成了感知、推理、学习和交互等多种能力，能够在复杂的环境中独立完成任务，或与人类进行协作。

2.1.2 AI 智能体的分类

按照不同的维度，AI 智能体可以分为多种类型。从任务类型维度划分，有简单反射智能体、基于模型反射智能体、基于目标智能体和基于效用智能体。简单反射智能体仅根据当前感知到的信息做出反应，不考虑历史数据和目标，如简单的温度控制系统，当温度高于设定值时，自动开启制冷设备，反之则关闭，它反应迅速，但在复杂多变的环境中，应对能力有限。

基于模型反射智能体不仅能感知当前环境，还能结合历史数据构建环境模型，从而更准确地做出决策。智能交通系统通过分析过往交通流量数据以及实时路况，来调整信号灯时长，以缓解交通拥堵。

基于目标智能体以实现特定目标为导向，利用知识和搜索算法寻找达成目标的最佳路径。机器人在复杂地形中执行探索任务时，会根据目标地点，规划出一条最优路线，避开障碍物，顺利抵达目的地。

基于效用智能体在决策时，会综合考虑各种因素，根据效用函数评估不同方案的价值，选择能带来最大效用的行动。在投资决策中，智能体可以分析市场行情、风险因素等，给出最优的投资组合建议。

从智能体的结构和功能维度划分，有单智能体和多智能体。单智能体是由一个独立系统运行，主要用于单一任务的高效执行，其特点是聚焦于某一特定领域，运行简单，开发和维护成本较低，适合特定的高频使用场景，如专门用于图像识别的智能体。多智能体系统则由多个智能体协同工作，通常应用于复杂、跨领域的任务，这些智能体可以分工负责不同任务，

通过信息共享完成整体目标。在物流配送中，多个智能体分别负责订单处理、车辆调度、路径规划等任务，共同完成货物的高效配送。

按照智能体的学习能力划分，有学习智能体和非学习智能体。学习智能体具备强大的学习能力，能从经验中不断学习，优化自身性能。智能客服通过与大量用户的交互，学习常见问题的解答方式，不断提升服务质量。非学习智能体则基于预先设定的规则和算法运行，不具备从经验中学习的能力。传统的基于规则的专家系统，在运行过程中不会根据新的情况更新自身的知识和策略。

2.2 大模型与 AI 智能体的融合机制

大模型作为 AI 智能体的“大脑”，为智能体提供了强大的语言交互、知识推理和决策能力。在语言交互方面，大模型能够理解自然语言，将用户的指令转化为可执行的任务。当用户向智能体提出问题或请求时，大模型通过自然语言处理技术解析用户的输入，理解用户的意图。在知识推理方面，大模型拥有丰富的知识储备，能够根据已知的信息进行推理和判断。智能体在面对复杂问题时，大模型可以利用其知识体系进行分析和推理，为智能体提供决策依据。当智能体需要解决一个技术难题时，大模型可以搜索相关的知识和解决方案，帮助智能体做出正确的决策。在决策方面，大模型能够根据智能体感知到的环境信息和任务目标，制定合理的行动策略。结合强化学习等技术，大模型可以不断优化智能体的决策过程，使其在动态环境中能够做出最优的决策。

为了实现大模型与 AI 智能体的有效融合，通常还需要结合其他技术。强化学习可以让智能体在与环境的交互中不断学习和优化自身的行为策略，通过奖励反馈机制，智能体能够逐步找到最优的行动方案。将强化学习与大模型相结合，大模型可以为智能体提供初始的策略和知识，智能体在实际运行中通过强化学习不断改进策略，提高决策能力。记忆模块也是融合机制中的重要组成部分，它可以帮助智能体存储和检索历史信息，包括过去的决策、环境状态和经验教训等。工具使用能力的赋予也是大模型与 AI 智能体融合的关键。智能体可以通过调用外部工具，如搜索引擎、数据库、计算器等，来获取更多的信息和资源，增强自身的能力。通过这些融合机制，基于大模型的 AI 智能体能够在复杂的环境中实现更智能、高效的行为。

三、AI 智能体的构建技术

3.1 构建框架与关键组件

3.1.1 智能体框架设计原则

通用性是智能体框架设计的重要原则之一，旨在使智能体能够适应多种不同类型的任务和场景。

可扩展性是智能体框架能够方便地添加新的功能模块、算法和数据，以满足不断变化的需求。

灵活性原则要求智能体框架能够根据不同的任务需求和环境条件，灵活地调整自身的行为和策略。

3.1.2 脑模块（Brain）

脑模块是 AI 智能体的核心组件，承担着自然语言交互、知识与记忆、推理与规划等重要功能。在自然语言交互方面，脑模块基于大语言模型，通过自然语言处理技术，实现与用户的流畅对话。当用户输入自然语言指令时，脑模块能够准确解析指令的语义和意图。

知识与记忆是脑模块的重要功能之一。大语言模型在预训练过程中，学习了海量的文本数据，积累了丰富的知识，包括常识知识、专业领域知识等。脑模块可以利用这些知识回答用户的问题，解决各种任务。在智能客服场景中，脑模块可以记录用户的问题和之前的解决方案，当用户再次询问类似问题时，能够快速给出答案，提高服务效率；在任务执行过程中，脑模块可以记录任务的中间状态和结果，为后续的决策和行动提供参考。

推理与规划能力是脑模块实现复杂任务的关键。对于复杂的逻辑问题，脑模块能够通过推理能力，分析问题的条件和要求，找到解决问题的方法。在数学问题求解中，脑模块可以运用逻辑推理和数学知识，逐步推导得出答案。对于多步任务，脑模块能够制定详细的规划，将任务分解为多个子任务，并合理安排子任务的执行顺序。

3.1.3 感知模块（Perception）

感知模块负责 AI 智能体对环境信息的获取和理解，实现文本、视觉、音频等多模态感知。在文本感知方面，感知模块利用自然语言处理技术，对输入的文本进行分析和理解。感知模块还能理解文本中的情感倾向，判断文本是积极、消极还是中性的情感表达。

视觉感知使智能体能够理解图像和视频等视觉内容。通过计算机视觉技术，感知模块可以对图像进行分析和处理。在图像识别任务中，它能够识别图像中的物体类别，判断图像中是人物、动物、风景还是其他物体。对于视频数据，感知模块可以分析视频的帧序列，识别视频中的动作和事件。

音频感知是感知模块处理声音和语音信号的能力。在语音识别方面，感知模块能够将语音信号转换为文本，实现人机语音交互。智能语音助手通过音频感知模块，接收用户的语音指令，将其转换为文本后传递给脑模块进行处理。感知模块还能分析音频中的环境声音，识别不同的声音类别。在智能家居系统中，音频感知模块可以根据环境声音的变化，自动调整设备的运行状态。

3.1.4 行动模块（Action）

行动模块是 AI 智能体将决策转化为实际行动的关键组件，具有文本输出、工具使用、实体化行动等功能。文本输出是行动模块与用户交互的基本方式之一。根据脑模块的决策和分析结果，行动模块生成自然语言文本进行回复。在智能问答系统中，当脑模块解答了用户的问题后，行动模块将答案以清晰、易懂的文本形式输出给用户。

工具使用能力拓展了智能体的功能和应用范围。智能体可以通过行动模块调用各种外部工具，以完成特定的任务。在信息检索任务中，行动模块可以调用搜索引擎工具，输入关键词，获取相关的信息和知识。

实体化行动使智能体能够与物理世界进行交互，控制物理设备执行具体的操作。在智能家居系统中，行动模块可以控制家电设备的运行。根据用户的指令或环境感知信息，行动模块发送控制信号，打开或关闭灯光、调节空调温度、启动洗衣机等。在工业生产领域，行动模块可以控制机器人执行生产任务。通过实体化行动，智能体能够在现实世界中发挥实际作用，解决各种实际问题。

3.2 工具集成与能力拓展

3.2.1 常用工具类型与作用

搜索工具是智能体获取外部信息的重要手段，如谷歌、百度等通用搜索引擎，以及专门针对学术领域的知网、万方等搜索平台。在信息查询任务中，智能体利用搜索工具，输入关键词，能够快速获取大量相关的文本、图片、视频等信息。

文件操作工具用于对文件进行管理和处理，包括文件的创建、读取、写入、删除、复制、移动等操作。在办公场景中，智能体可以利用文件操作工具对文档、表格、演示文稿等文件进行处理。

数据分析工具在处理大量数据时发挥着关键作用，常见的有 Python 的数据分析库（如 Pandas、NumPy、Matplotlib 等）、专业的数据分析软件（如 SPSS、SAS 等）。在商业领域，智能体使用数据分析工具对销售数据、市场调研数据等进行分析。

3.2.2 工具集成方式

工具集成的方式主要有 API 调用和插件式集成。API 调用是智能体通过调用工具提供的应用程序编程接口（API）来使用工具的功能。许多搜索引擎、地图服务、云存储等工具

都提供了 API，智能体可以通过发送 HTTP 请求，按照 API 的规范传递参数，获取工具返回的结果。

插件式集成是将工具以插件的形式集成到智能体系统中，插件与智能体系统之间通过特定的接口进行通信和交互。插件式集成可以实现更紧密的结合，智能体能够更好地控制和管理工具的使用。

3.3 多智能体系统构建

3.3.1 多智能体系统的架构模式

集中式架构模式下，存在一个中央控制单元，负责管理和协调所有智能体的行为。中央控制单元收集各个智能体的感知信息，进行统一的决策和规划，然后向智能体发送指令，指导它们的行动。

分布式架构模式中，各个智能体之间相互独立，通过通信和协作来完成共同的任务。每个智能体都有自己的感知、决策和行动能力，它们根据自身的目标和环境信息，自主地做出决策，并与其他智能体进行信息共享和交互。

混合式架构模式结合了集中式和分布式架构的优点，既有中央控制单元进行全局的协调和管理，又允许智能体在一定程度上自主决策和行动。混合式架构的设计和实现较为复杂，需要在集中控制和分布式自主之间找到合适的平衡点。

四、创新场景应用探索

4.1 行业应用案例深度解析

4.1.1 医疗领域：智能诊断与健康监测

在疾病诊断方面，AI 智能体展现出了卓越的能力。通过对海量医疗数据的学习，智能体能够快速准确地分析患者的症状、病史和检查结果，辅助医生做出更精准的诊断。以影像诊断为例，AI 智能体利用深度学习算法，能够对 X 光、CT、MRI 等医学影像进行快速分析，识别出病变部位和特征，为医生提供诊断建议。在实际应用中，AI 智能体可以在短时间内处理大量的影像数据，帮助医生快速筛选出疑似病例，提高诊断效率。

在健康监测方面，AI 智能体借助可穿戴设备和传感器技术，实现对用户健康状况的实时跟踪。智能手环、智能手表等可穿戴设备可以收集用户的心率、血压、睡眠质量、运动步数等生理数据，并将这些数据传输给 AI 智能体。智能体通过数据分析和机器学习算法，建立用户的健康模型，预测疾病风险，提供个性化的健康建议。通过长期的健康数据监测，智能体还可以根据用户的健康状况和生活习惯，制定个性化的运动和饮食计划，帮助用户预防疾病，保持健康。

4.1.2 金融领域：风险评估与投资决策

在金融风险评估方面，AI 智能体能够综合分析大量的金融数据，包括市场行情、企业财务报表、信用记录等，对金融风险进行精准评估。AI 智能体利用机器学习和深度学习算法，能够挖掘数据中的潜在模式和关联，建立更加准确的风险评估模型。在信用风险评估中，智能体可以通过分析借款人的个人信息、历史还款记录、消费行为等多维度数据，预测其违约风险，为金融机构的信贷决策提供依据。通过对市场数据的实时监测和分析，智能体还能及时发现市场风险的变化，帮助金融机构提前采取风险防范措施。

在投资决策方面，AI 智能体可以根据市场动态和投资者的风险偏好，制定个性化的投资策略。智能体通过对海量金融数据的分析，预测资产价格的走势，寻找投资机会。它可以实时跟踪股票、债券、基金等各类资产的市场表现，结合宏观经济数据和行业发展趋势，为投资者提供投资建议。智能体还能根据投资者的风险承受能力和投资目标，优化资产配置，实现风险与收益的平衡。通过持续学习和实时更新市场信息，AI 智能体能够不断优化投资策略，提高投资回报率。

4.1.3 教育领域：个性化学习与辅导

在个性化学习路径规划方面, AI 智能体可以根据学生的学习能力、知识水平、学习风格和兴趣爱好等多方面因素, 为每个学生制定专属的学习计划。通过对学生学习数据的分析, 智能体能够了解学生的学习进度和掌握情况, 发现学生的学习难点和薄弱环节。教师通过在线测试和课堂观察收集学生的学习数据, 智能体根据这些数据生成个性化的学习材料, 包括针对性的练习题、知识点讲解视频等。随着学生学习过程的推进, 智能体还能根据学生的实时学习情况动态调整学习计划, 确保学习计划始终符合学生的需求。

在智能辅导方面, AI 智能体能够实时解答学生的问题, 提供一对一的辅导服务。当学生在学习过程中遇到问题时, 可以向智能体提问, 智能体通过自然语言处理技术理解学生的问题, 并运用其知识储备和推理能力给出准确的解答。智能体不仅能回答学生的知识性问题, 还能引导学生进行思考, 帮助学生掌握解题思路和方法。智能体还可以通过与学生的互动, 了解学生的学习状态和情绪, 给予鼓励和支持, 提高学生的学习积极性。

4.1.4 制造业: 生产优化与质量控制

在生产流程优化方面, AI 智能体可以实时监测生产线上的各种数据, 包括设备运行状态、生产进度、原材料消耗等, 通过数据分析和机器学习算法, 找出生产流程中的瓶颈和优化空间。智能体可以根据订单需求、设备产能和原材料供应情况, 优化生产调度, 合理安排生产任务, 提高生产效率。

在质量检测方面, AI 智能体利用计算机视觉和机器学习技术, 对产品进行快速、准确的质量检测。传统的人工质量检测方式不仅效率低下, 而且容易受到人为因素的影响, 导致检测结果不准确。AI 智能体可以通过摄像头采集产品的图像信息, 分析产品的外观、尺寸、形状等特征, 判断产品是否符合质量标准。

4.2 新兴应用场景展望

4.2.1 智能家居与智能城市

在智能家居方面, AI 智能体可以实现对家居设备的全面智能控制。用户只需通过语音指令, 智能体就能控制灯光、空调、电视、窗帘等设备。通过对用户行为数据的学习, 智能体还能提供个性化的服务。在家庭安全方面, 智能体可以通过智能摄像头、智能门锁等设备, 实时监测家庭安全状况。

在智能城市中, AI 智能体在城市交通管理和能源优化方面具有巨大的应用潜力。在交通管理方面, 智能体可以实时收集交通流量、车辆位置、路况等信息, 通过数据分析和算法优化, 实现交通信号灯的智能配时, 缓解交通拥堵。智能体还能为车辆提供实时的导航建议, 根据路况动态规划最优行驶路线, 减少车辆行驶时间和能源消耗。

4.2.2 娱乐与文化创意产业

在游戏领域, AI 智能体可以作为游戏中的虚拟角色, 与玩家进行互动, 提升游戏的趣味性和挑战性。AI 智能体能够根据玩家的游戏风格 and 行为习惯, 动态调整游戏策略, 提供个性化的游戏体验。

在影视创作方面, AI 智能体可以辅助编剧进行剧本创作。通过对大量影视作品和文学作品的学习, 智能体能够理解故事结构、人物塑造、情节发展等元素, 为编剧提供创意和灵感。

六、结论

本研究深入剖析了基于大模型的 AI 智能体构建技术, 从理论基础到实际应用。在理论层面, 系统阐述了 AI 智能体的基本概念、特性与分类, 以及大模型与 AI 智能体之间的融合机制。这为后续的技术研究和应用开发提供了坚实的理论支撑, 使得我们能够从本质上理解 AI 智能体的智能实现方式, 以及大模型在其中所发挥的关键作用。

在构建技术方面, 设计了通用、可扩展且灵活的智能体框架, 详细阐述了脑模块、感知模块和行动模块等关键组件的功能和实现方式。通过工具集成, 拓展了智能体的能力, 使其

能够借助搜索工具、文件操作工具、数据分析工具等完成更复杂的任务。在多智能体系统构建方面,研究了集中式、分布式和混合式等架构模式,以及智能体间的协作与通信机制,为解决复杂任务提供了有效的途径。

在创新场景应用探索中,对医疗、金融、教育、制造业等行业进行了深度案例分析,展示了 AI 智能体在这些领域的强大应用潜力。还对智能家居、智能城市、娱乐与文化创意产业等新兴应用场景进行了展望,为 AI 智能体的未来应用拓展提供了方向。

未来的研究工作将围绕以下方面展开。在技术研究方面,将致力于探索更高效的模型架构和训练方法,降低大模型的计算资源需求,提高模型的推理准确性和可解释性。研究新的模型架构,开发新的训练算法,使模型能够更快地学习和适应新任务。在伦理与安全方面,将制定完善的伦理规范和法律法规,加强对 AI 智能体的数据隐私保护和算法监管,防范恶意攻击。建立严格的数据隐私保护机制,采用加密技术和访问控制等手段,确保用户数据的安全。还将继续探索 AI 智能体在更多领域的创新应用,推动人工智能技术的广泛应用和发展。

本论文得到 2024 年中国人生科学学会“十四五”科研课题“基于大模型的 AI 智能体的构建与创新场景应用探索”项目资助,课题编号:K1102024061236。

参考文献

- [1] 周涛,李鑫,周俊临,等.大模型智能体:概念、前沿和产业实践[J].电子科技大学学报(社科版),2024,26(04):57-62.
- [2] 刘明,杨闽,吴忠明,等.教育大模型智能体的开发、应用现状与未来展望[J].现代教育技术,2024,34(11):5-14.
- [3] 刘阳.智能体交互模式大模型时代的新型交互范式[J].通信世界,2024,(13):26-29.
- [4] 卢宇,余京蕾,陈鹏鹤.基于大模型的教学智能体构建与应用研究[J].中国电化教育,2024,(07):99-108.
- [5] 王明程,王高开,李勇男.基于大模型智能体的安全风险态势感知框架构建[J].情报理论与实践,2024,47(07):190-198.

Construction and Innovative Scenario Application Exploration of large language model-Based AI Agents

Liu Xiaoyu

(College of Applied Science and Technology, Beijing Union University, Beijing, 100012)

Abstract: This paper focuses on the construction of large language model (LLM)-based AI agents and their innovative application scenarios. By reviewing current research progress globally, it showcases cutting-edge explorations by international institutions (e.g., OpenAI, Google, DeepMind) and domestic organizations (e.g., Tsinghua University, Peking University, and Chinese enterprises) in this field. The study systematically analyzes fundamental theories of AI agents, including definitions, characteristics, classification criteria, and their integration mechanisms with LLM technology. It thoroughly investigates agent construction methodologies, covering architectural frameworks, core components, tool integration strategies, and multi-agent system development. Through practical case studies across healthcare, finance, education, and manufacturing industries, the paper conducts in-depth analysis of agent application patterns and value propositions, while exploring prospects in emerging scenarios like smart homes and entertainment. Finally, it identifies current research challenges, proposes targeted solutions and future development directions, aiming to provide theoretical and practical references for advancing LLM-based AI agent technologies and their widespread adoption.

Keywords: Large Models; AI Agents; Construction Methodologies; Innovative Scenarios; Industry Cases