

智能体技术和应用研究报告

(2025 年)

中国信息通信研究院人工智能研究所

华为技术有限公司

2025年6月

版权声明

本报告版权属于中国信息通信研究院和华为技术有限公司，并受法律保护。转载、摘编或利用其它方式使用本报告文字或者观点的，应注明“来源：中国信息通信研究院和华为技术有限公司”。违反上述声明者，编者将追究其相关法律责任。

前 言

近年来，以大模型为代表的新一代人工智能技术迎来爆发式增长，成为推动产业升级、促进经济发展和引领社会进步的重要力量。智能体作为大模型的原生应用形态之一，能将模型各类能力转化为任务执行能力，高度贴合日益复杂的提质增效需求，加速推动行业数字化转型和智能化升级。通过梳理智能体技术和应用发展现状，呈现行业应用成效，分析当前问题挑战，可为智能体产业生态建设提供系统性参考。

本报告重点围绕智能体发展概述、关键技术、产业应用、问题挑战、发展建议五大方面，深刻剖析当前智能体发展现状及核心问题。**发展概述层面**，梳理了智能体的发展脉络，论述了智能体的概念演进、技术及应用优势，分析了智能体的社会价值。**关键技术层面**，聚焦模型多维能力、“脑手”双轮驱动、通信协议以及多智能体协同等方面，深度分析了智能体的关键技术进展，推动技术创新融合。**产业应用层面**，阐述了智能体的平台工具应用、行业场景赋能、生态体系建设等情况，给出了智能体行业实践案例，为产学研用各方提供有效参考，加速智能体落地应用进程。**问题挑战层面**，论述了智能体在认知规划协同能力、应用场景创新及个性化、安全伦理及隐私保护等方面的挑战。**发展建议层面**，围绕夯实智能技术底座、促进多领域落地应用以及对齐人类价值偏好等维度进行了展望，为智能体的向上向阳发展提供了有力支撑。

未来，加快推进智能体技术创新和产业应用将成为推进人工智能

与实体经济深度融合的重要举措。聚焦智能体核心技术研发的深度布局，加速构建智能体应用体系，有助于高效培育新质生产力，促进智能体普惠化、可持续、高质量发展。当前智能体技术正处于快速发展阶段，智能体技术和应用研究工作还将持续深化开展，报告中如有不足之处还请各方专家读者不吝指正。

目 录

一、智能体发展概述.....	1
（一）兼具科研应用双重价值，初显通用智能雏形	1
（二）破解模型有脑无手困局，释放智能应用潜力	5
（三）培育和发展新质生产力，开启原生应用时代	7
二、智能体关键技术.....	9
（一）模型多维技术创新融合，提升感知认知能力	10
（二）全局规划最优工作流程，记忆多元历史信息	12
（三）工具调用拓宽能力范围，高效执行复杂任务	14
（四）通信协议夯实原生基石，加固多体协同底座	17
三、智能体产业应用.....	20
（一）平台工具加速落地进程，构建全域生态体系	21
（二）通用场景实现百花齐放，助力实现降本增效	26
（三）专用场景持续走深向实，加速推进数转智改	30
四、智能体问题挑战.....	34
（一）技术底座能力发展放缓，认知规划能力不足	34
（二）应用场景创新面临挑战，个性需求挖掘不深	37
（三）安全治理体系仍不完善，标准规范构建不全	39
五、智能体发展建议.....	41
（一）全方位加强大模型攻关，夯实智能技术底座	41
（二）多领域促进智能体落地，推动应用普惠向实	43
（三）高标准引导可信赖发展，对齐人类价值偏好	44

图 目 录

图 1 华为 AI 原生应用引擎平台	9
图 2 智能体核心能力图	10
图 3 华为 Data Agent 总体架构图	16
图 4 故障监控专家智能体	20
图 5 智能体基础支撑应用图谱	21
图 6 蚂蚁智能体开发平台	23
图 7 “交心”智能体平台架构	25
图 8 智能体通用场景应用图谱	27
图 9 华为 Operator 智能体总体架构图	28
图 10 千帆 AppBuilder 智能体平台	30
图 11 智能体专用场景应用图谱	31
图 12 华为 Deep Research 总体架构图	32
图 13 蝴蝶大模型智能体服务平台	34

一、智能体发展概述

智能体（AI Agent）作为能自主设定和实现目标的代理，其演进之路伴随着人工智能（AI）技术的持续发展，实现从自动化工具到自主性能力的升级。20 世纪 50 年代，博弈论、控制论、信息论及图灵测试等相继为智能体的研究提供了理论基础。1986 年，马文·明斯基首次在 AI 领域引入智能体概念¹。20 世纪 90 年代，随着机器学习和自然语言处理技术的突破，智能体的研究扩展至多模态感知交互和自主学习，智能体的定义演化成了能感知环境、影响环境的实体。

（一）兼具科研应用双重价值，初显通用智能雏形

1. 智能体概念定义及国内外现状

智能体是指能够感知环境并利用工具采取行动以实现特定目标的代理。智能体以大模型为智能底座，具备自主感知、理解、规划、决策、记忆、行动和使用工具的能力，能够自动化执行复杂任务，具有自主性、交互性、反应性和适应性等基本特征。

目前，智能体技术特点初步显现，应用场景逐渐深化，国内外智能体产品加速涌现。智能体已渗透至电信、制造、金融、政务、能源、互联网等垂直行业，涵盖数据分析、生产管理、智能客服、智能运维等典型场景。根据 Marketsand Markets 预测²，全球智能体市场将从 2024 年的 51 亿美元增长到 2030 年的 471 亿美元，年复合增长率达 44.8%。国外智能体产品及市场现状上，智能体技术发展如火如荼，

¹ Minsky, Marvin. The Society of Mind. Simon & Schuster, 1986.

² <https://www.marketsandmarkets.com/Market-Reports/ai-agents-market-15761548.html>

Google、Amazon、OpenAI、Anthropic 等均已推出智能体相关产品，如 Google Agentspace、Amazon Bedrock Agents、OpenAI Operator、OpenAI Deep Research、Claude Code 等。国内智能体产品及市场现状上，智能体技术创新活跃，产品百花齐放。国内企业积极在智能体的生态拓展和应用市场开拓方面持续探索，开发出 Data Agent、Operator Agent、扣子、MiniMax Agent、Manus、AppBuilder、Agentar 等多种产品。各大科技公司都在快速迭代智能体产品与解决方案。

国内外智能体在技术演进、产品创新、场景应用、生态建设等方面各具特点。在技术演进方面，国内外均高度重视智能体技术研发和应用，国外聚焦理论和技术创新，我国侧重技术落地和实践。国外智能体开发注重前沿技术创新与通用能力建设，在智能体理论研究、技术架构等方面具有领先优势，主导通用智能体技术创新。我国在智能体轻量化、环境适应、工具调用等技术方面取得进展。在产品创新方面，国内外均基于大模型能力构建智能体平台产品。国外积极布局一体化端侧智能体产品的开发，我国深入探索专用型、定制化的智能体产品，提供定制化、专业化的解决方案。在场景应用方面，国内外均积极推广智能体行业应用，持续探索商业模式。国外深化探索金融、医疗、电商等垂直行业应用落地需求。我国以场景驱动快速切入，智能体应用更注重客服、营销、助手等场景，并初步取得成效。在生态建设方面，国内外均在持续完善智能体生态体系建设。国外通过算法创新、开源开放等技术手段驱动生态完善。我国主要聚焦政府、市场

双轮驱动促进生态发展。Deloitte 预测³，到 2025 年，将有 25% 的企业部署生成式 AI 驱动的智能代理，2027 年这一比例将升至 50%。

2. 人工智能及智能体发展的相关政策

各国政府纷纷出台相关政策和战略规划，以推动人工智能及智能体的发展。国际方面，欧盟通过《人工智能法案》，对高风险智能体实施严格准入审查。英国发布《生成式人工智能框架》，推动人工智能的负责任使用和创新。美国推出《人工智能创新未来法案》，强调了国际标准的制定、数据共享和安全性研究的重要性。国内方面，我国智能体发展深度嵌入“数字中国”战略，政策层面形成“中央定调-地方试点-行业标准”三级推进体系。国家政策以顶层战略为指引，多部门联合构建全面政策框架，统筹人工智能发展战略方向、技术路径与安全治理；地方政策以国家政策为导向，结合本地资源禀赋和产业基础，围绕产业发展关键环节出台配套措施。工业和信息化部、国家发展和改革委员会等四部门联合编制发布了《国家人工智能产业综合标准化体系建设指南（2024 版）》⁴，强调到 2026 年将新制定国家标准和行业标准 50 项以上，引领人工智能产业高质量发展的标准体系加快形成。将参与制定国际标准 20 项以上，促进人工智能产业全球化发展。

3. 智能体具备科研及应用价值

随着智能体技术的不断发展，其科研和应用价值逐渐清晰，为探索通用人工智能提供了关键路径。科研价值方面，智能体的科研价值

³ <https://www.deloitte.com/global/en/about/press-room/deloitte-globals-2025-predictions-report.html>

⁴ https://www.gov.cn/zhengce/zhengceku/202407/content_6960720.html

主要是作为人工智能研究的核心载体，能够推动基础理论创新并促进跨学科融合。智能体研究催生了新型强化学习算法和多智能体协作机制，并支持大规模并行实验和可重复验证，显著加速科研进程，并已在天文发现、材料研发等科学领域取得成果。**应用价值方面**，智能体的应用价值主要是其作为智能化解决方案的核心载体，能够显著提升各行业效率并创造新服务模式。智能体已在基础科学、金融、工业、教育、医疗、政务等行业得到应用，并赋能智慧办公、研发设计、生产制造、经营管理、客户服务等多种应用场景。通过持续的技术迭代，不断拓展人类能力的边界，推动社会向智能化时代加速迈进。根据 Gartner 预测⁵，到 2028 年，至少 15% 的日常工作决策将通过人工智能代理自主完成。

4. 智能体加速通用人工智能实现步伐

通用人工智能的实现是一个长期且复杂的过程，智能体被看作是实现通用人工智能的重要手段，主要体现在以下几个方面。**环境交互方面**，智能体能够通过不同传感器感知和理解周围环境的多模态信息，包括文本、声音、图像等，并能够以多种方式与环境、用户、其他智能体进行交互，实现更加智能、自然和人性化的交互体验。**持续学习方面**，智能体在与环境交互的过程中可以不断获取新的任务和数据，通过持续演进不断提高自身的性能和功能，逐步接近或超越人类的智能水平。**决策推理方面**，智能体能够根据感知到的信息和自身储备的知识，进行推理和决策，选择最优的行动方案。**动态适应方面**，智能

⁵ <https://www.gartner.com/en/newsroom/press-releases/2024-10-21-gartner-identifies-the-top-10-strategic-technology-trends-for-2025>

体能够在面对不同环境和任务时，灵活调整自己的行为 and 策略，以适应新的情况 and 要求。**自主规划方面**，智能体能够在一定程度上独立地进行思考、决策 and 行动，而不是完全受限于外部指令 or 固定程序。智能体能够通过复杂的规划和优化算法，自主地从经验中学习，并根据学习到的知识和经验高效执行新任务。

（二）破解模型有脑无手困局，释放智能应用潜力

智能体作为大模型的重要应用形态，与大模型相辅相成。智能体为大模型提供实际场景的具体应用，大模型为智能体提供强大的问题处理能力和运算能力，进而使得智能体在多种环境下完成复杂操作。智能体可以有效地克服大模型“有脑无手”的问题，助力大模型在实际场景的落地应用。

1. 大模型感知和认知能力显著提升

大模型在多维度表现出卓越的感知和认知能力，逐步接近人类水平。**感知能力方面**，大模型具有强大的单任务识别分类能力，能根据不同类别之间的特征表示准确识别出不同的对象。大模型具有强大的多模态融合能力，可凭借强大的表征和泛化能力同时处理大量文本、图像、视频、音频等不同模态的信息。**认知能力方面**，大模型具有强大的理解和生成能力，结合跨领域知识图谱、上下文学习等技术可实现复杂问题的理解，通过序列到序列模型、扩散模型等技术实现生成创造。大模型具有强大的推理和决策能力，可以通过思维链、提示学习、强化学习等技术挖掘数据之间的关联关系实现推理和决策。大模型具有自主学习和适应能力，可以根据环境和任务的变化自动调整自

身参数和结构。

2. 大模型“有脑无手”应用落地困难

当前，大模型已经成为 AI 行业的代名词，蓬勃发展的各类大模型赋能千行百业。尽管大模型在感知和认知能力方面获得了巨大提升，但却缺乏直接的行动能力。交互方面，大模型通常在服务器或云端中运行，缺乏物理接口和感知物理世界变化的传感器，不具备与物理世界直接交互的能力。执行方面，大模型本身无法执行物理动作，需要与机器人、自动化设备等外部系统集成才能实现动作执行。实时性方面，大模型通常需要较长的时间来进行推理并生成输出，不能适应需要快速响应的任务。可扩展性方面，大模型在扩展到新的任务时，通常依赖大量的数据和重新训练。

3. 智能体加速大模型应用落地

智能体进一步释放大模型潜力，将模型能力转化为实际问题的解决能力，打通业务场景和人类交互的桥梁。智能体构建框架方面，智能体框架具备高度的定制化和集成能力，极大地扩展了应用场景及可能性。多智能体协同方面，通过对任务的合理分解、共享知识、执行子任务和对齐目标，进而优化资源配置，显著提高对复杂任务的执行和问题解决能力。智能体 workflow 方面，通过合理的任务规划、多次迭代及优化技术，可以实现更高质量的输出，助力模型性能大幅提升，确保任务顺利高效完成。端侧智能方面，通过模型压缩和硬件加速实现性能与效率的平衡，减少对云端的依赖，提高响应速度和隐私保护能力。

（三）培育和发展新质生产力，开启原生应用时代

新质生产力源自技术的革命性突破，当前智能体已展现出对人工智能高质量发展的强劲推动力和支撑力。智能体将重塑人工智能创新生态，大幅提升应用方的生产效率。随着智能体能力的进一步提升，将催生出大量的 AI 原生应用，更好地服务复杂的业务场景。

1. 智能体赋能新质生产力发展

智能体作为人工智能领域的重要发展方向，有望成为新质生产力发展重要引擎。推动技术升级方面，智能体加速创新驱动，智能体的发展和应用离不开原创模型和算法的研发，通过不断的优化和迭代，智能体能够提供更加高效、准确和智能的解决方案。助力协同共享及全局优化方面，智能体与行业平台和大模型相结合，打破数据孤岛，实现全流程信息化覆盖。能够通过多智能体协同技术，实现复杂任务的自主决策和全局优化。优化生产流程和效率方面，智能体拓展更多应用场景，智能体通过其自主性、交互性和适应性，正在不断扩展人工智能的边界。智能体通过技术嵌入和流程重构等方式，推动企业从传统生产模式向智能化转型。促进可信赖发展方面，智能体的广泛应用需要对人工智能的治理和伦理问题进行深入探讨和规范，以确保人工智能应用的可信、可控和可解释性。

2. 智能体激发原生应用新活力

智能体通过“环境感知-自主决策-动态执行”的核心能力，为原生应用带来了新的活力，推动各行业的创新发展。一是智能体为 AI 原生应用提供强大的技术支撑。智能体通过自我反思和优化，不断提

升 AI 原生应用的性能和功能，助力 AI 原生应用实现智能化、自动化和高效化。**二是**自主决策能力突破了传统 AI 工具被动响应的局限，使其能够更好地适应复杂、动态的环境和多样化的用户需求，实现从“建议者”到“执行者”的转变。**三是**多智能体协作机制创造了跨领域融合创新的可能，智能体的持续进化特性使应用具备自我迭代能力，不仅重构了人机协作模式，更在金融、制造等领域培育出“智能体经济”新生态。

华为 AI 原生应用引擎

华为 AI 原生应用引擎是企业专属的一站式大模型开发及应用构建平台。无论是研发技术人员还是业务人员，都可通过简易的界面化操作，快速开发大模型应用或训练专属模型。

AI 原生应用引擎提供企业专属大模型开发和应用开发的整套工具链和关键技术栈，包括主流大模型内置、数据准备、模型调优、知识工程、智能模式/工作流模式智能体开发、应用集成等能力，降低智能应用开发门槛、提升开发效率，助力企业客户将专属大模型能力融入自己的业务应用链路或对外应用服务中；同时提供 AI 资产中心，对接并承载内外部伙伴行业 AI 资产，形成行业 Know-how 对外展现与复制。

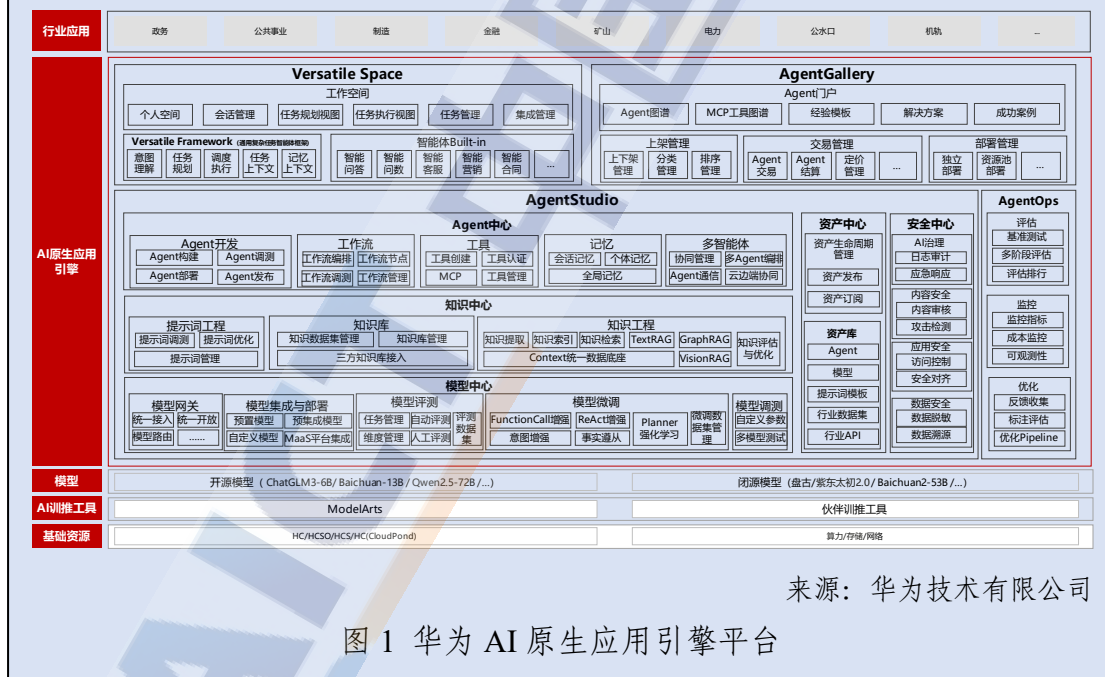
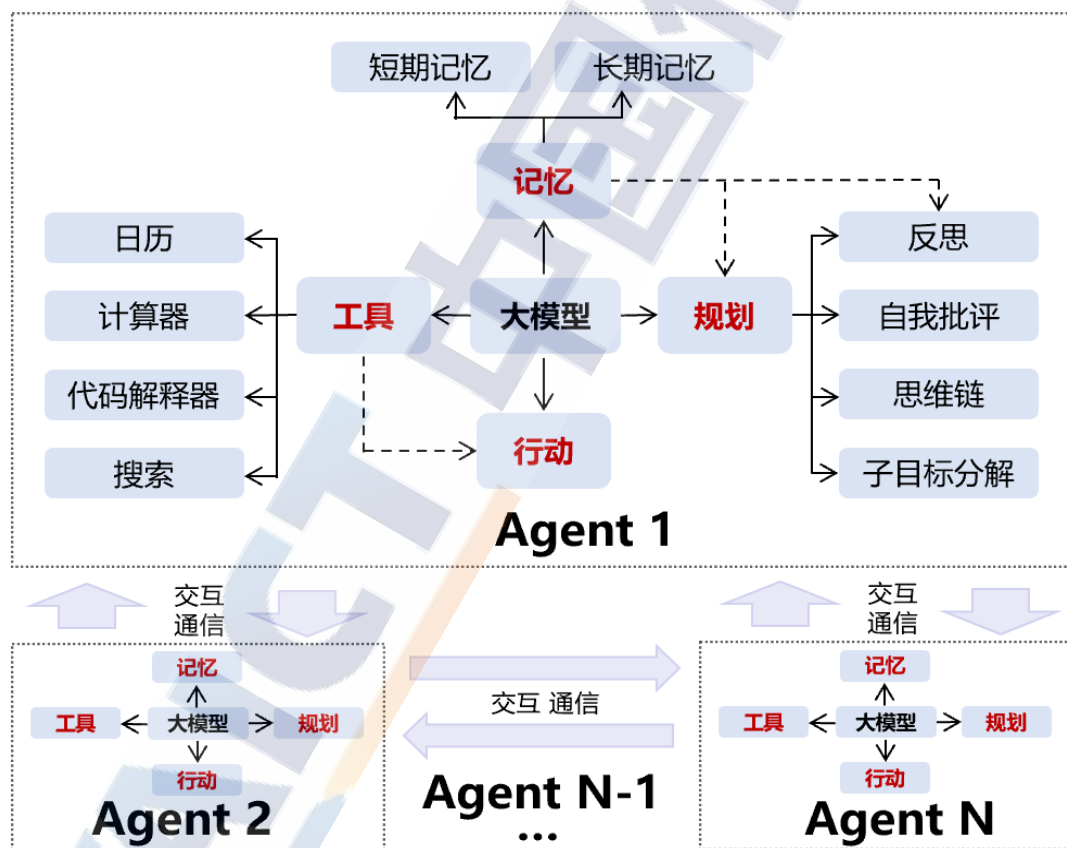


图 1 华为 AI 原生应用引擎平台

二、智能体关键技术

大模型作为智能体的“大脑”，让智能体具备更强大的学习、生成、推理、决策能力。通过“脑”侧的规划、记忆与思考机制，支撑

复杂任务的理解、分解与决策生成；结合“手”侧的工具调用与指令执行能力，确保任务方案在真实环境中高效落地。通过多个智能体协同可以同时实现任务处理质量和效率的提升，以 A2A、ANP、MCP 为代表的大模型和智能体通信协议的发展，为信息孤岛和通信兼容性问题提供了有效的技术解决方案，不仅助力大模型和智能体高效集成外部数据资源、工具或服务以拓展能力边界，还满足行业对跨场景、跨领域、跨层级的智能体协同需求，推动人工智能技术在复杂环境中的落地应用。



来源：公开资料整理

图 2 智能体核心能力图

（一）模型多维技术创新融合，提升感知认知能力

通过融合大模型的多维技术能力，包括多种模态融合、大小模型

协同等，智能体能够更高效地处理复杂任务，并结合人机协作机制，显著提升其在动态环境中的表现并拓展了智能应用的边界，为行业智能化转型提供了坚实的技术支撑。

1. 多模态大模型拓展智能体感知边界

多模态大模型通过整合文本、图像、音频和视频等信息，帮助智能体更好地理解环境，显著提升智能体的感知能力和智能水平。一是**跨模态理解**，智能体能够处理和关联不同形式的的数据，实现跨模态的语义映射与信息互补，从而增强对复杂信息的综合理解能力。二是**多模态交互**，智能体可以通过多种感官渠道与用户交互，通过整合多种输入和输出方式，提升用户体验，使交互更加自然和高效。三是**综合决策**，智能体能够基于多种模态的数据进行综合分析和决策，可有效提高决策效率、准确性和灵活性，并适应复杂多变的环境。

2. 检索增强生成丰富智能体知识储备边界

检索增强生成（RAG）技术辅助智能体提升信息存储利用率，提高信息更新速度，增强内容生成的可追溯性，并支持智能体的持续学习和迭代。RAG 技术结合了信息检索和生成模型，通过从外部知识库中检索相关信息来辅助生成更准确和上下文相关的文本。Agentic RAG 则可以通过迭代推理主动优化搜索，实现更准确、可解释和适应性强的响应。在**信息存储方面**，RAG 技术利用自然语言文本、数值向量、结构化数据、非结构化数据以及多模态数据，并利用关系型数据库、向量数据库等存储方式，为智能体提供了可查询的知识管理系统。在**信息管理方面**，通过实时更新外部知识库，RAG 技术提高输出

内容的及时性和准确性，显著提升了知识更新速度。用户可以查看生成内容的来源，验证其准确性和可靠性，增强了内容生成的可追溯性，提高了用户对智能体生成内容的信任度。在信息检索和生成方面，RAG 技术通过评估知识的更新时间、相关性和重要性，确定检索优先级，确保调用最相关的信息。生成模块则利用检索到的上下文信息，结合用户的查询，生成准确且上下文相关的响应。

3. 大小模型协同提升智能体任务适应边界

大小模型协同工作模式通过取长补短，显著提升了智能体的任务适应能力。大模型性能强大，但计算和存储资源消耗巨大，难以在资源受限场景应用；小模型轻便，性能却有限，协同工作则充分发挥两者优势。在任务分配方面，根据任务复杂度和资源需求，智能选择大模型或小模型。例如，深度思考任务使用大模型，快速响应任务使用小模型。模型压缩与知识蒸馏方面，通过将大模型知识和能力迁移到小模型，使其在资源受限环境下仍保持较高性能。混合推理方面，结合大模型和小模型推理结果，通过投票或加权平均等方式生成最终答案，兼顾性能与效率。大小模型协同方面，可以采用上下文构建和任务分解等方式，在保证智能体性能的同时，显著降低资源消耗，提高应用的可部署性和可扩展性。

（二）全局规划最优工作流程，记忆多元历史信息

智能体需具备灵活的规划能力与较强的记忆能力，规划能力指导智能体通过最优行动路径来完成任务，记忆能力帮助智能体积累和利用历史信息，通过整合短期与长期记忆提升任务执行的准确性和效率。

1. 规划能力制定最优行动路径

智能体的规划能力使智能体具备在复杂环境中为实现特定目标制定最优行动策略的能力。任务分解和 CoT 可以帮助智能体更好地规划复杂任务。智能体在处理复杂任务时，通过任务分解策略，将复杂任务分解为若干较小、易于处理的子任务，然后逐个解决，使问题的复杂度大大降低。ReAct 和 Reflexion 提升智能体的自主规划能力⁶。ReAct 强调推理和反思的重要性，让智能体能够在多变的环境中，透过理解任务、推理规划、执行行动和反思调整的循环过程，从而做出有效决策。Reflexion 技术则让智能体在面对新任务时，做出更合理的决策，并通过动态记忆和自我反思机制，使学习错误来优化自己的行动计划。LLM+P 和 ToT 方法解决长期规划问题⁷。LLM+P 方法是结合 LLM 的自然语言处理能力和传统规划方法两方面的优势，从而解决长期规划问题。ToT 方法为了有效提升智能体在多步骤推理任务中的表现，需要首先构建思维树，将复杂问题分解为多个子问题，对每个子问题分别求解，最后合并答案输出结果。CoH 技术提高人类价值观对齐程度⁸。CoH 技术通过将反馈信息序列化处理并用于模型微调，提高智能体与人类偏好的对齐程度，智能体可以自动学习用户的反馈意见，持续改进自身性能。

2. 记忆能力补充历史沟通记录

记忆能力高效存储、管理并检索信息，以满足不同应用场景的需

⁶ <https://research.google/blog/react-synergizing-reasoning-and-acting-in-language-models/>

⁷ <https://arxiv.org/pdf/2402.02716>

⁸ <https://arxiv.org/pdf/2302.02676>

求。在智能体中，记忆功能遵循由感觉记忆到短期记忆，进而转化为长期记忆的过程。智能体的短期记忆涉及当前情境的最新感知信息，而长期记忆则积累了智能体的经验和知识，形成历史数据。长期记忆对于智能体进行有效的推理和规划极为关键。多元记忆储存技术优化智能体的信息检索效能。在涉及完整语境和明确意图的场景下，自然语言格式的记忆存储表现出其适用性。而面对大规模数据和快速检索需求时，智能体则采用嵌入式技术，将信息转化为向量形式并通过向量数据库高效存储和检索，能迅速找到与当前任务最相关的信息。结构化的数据库则为智能体提供了一套可查询的记忆管理能力，允许实现对记忆内容的精确控制和查询。记忆管理技术融合 Reflexion 技术持续提升学习迭代能力⁹。引入 Reflexion 技术后，智能体通过文本反馈动态管理记忆，显著增强在序列任务中的决策性能，使智能体能够在连续决策和长期记忆的场景中精准响应，持续学习和调整迭代。

（三）工具调用拓宽能力范围，高效执行复杂任务

工具调用能力使智能体能够突破模型限制，通过 MCP 等多种方式调用外部资源，实现更广泛的互动和功能扩展。执行能力确保规划的目标能够落地为具体操作，通过多步骤实现任务的高效完成。智能体规划、记忆、工具和执行相互协作，共同构建起智能体从感知到执行的完整闭环系统。

1. 工具调用提升外界互动能力

工具调用助力智能体与外界互动能力提升，实现更高效的数据处

⁹ Shinn N, Cassano F, Gopinath A, et al. Reflexion: Language agents with verbal reinforcement learning[J]. Advances in Neural Information Processing Systems, 2023, 36: 8634-8652.

理及感知交互。智能体可以使用组件、知识库和数据库等多种类型工具，使智能体能够适应不同的业务需求，实现更加复杂和多样化的功能，为用户带来更多的便利和价值。**工具调用对智能体的数据分析处理能力起着举足轻重的作用。**SQL 和 NoSQL 数据库系统为智能体提供了数据存储和查询的能力。智能体的决策支持系统能够利用 Apache Spark 和 Hadoop, 对大规模数据集进行数据挖掘与分析处理，以揭示复杂数据模式。此外，智能体还可利用语音识别和合成技术，如 whisper 和 Festival 对语音数据进行加工处理。**工具调用加强智能体对周围的实时感知和监控。**通过 MQTT 协议和 IoT 平台等物联网工具，智能体能够借助摄像头、麦克风等传感器，实现对周围环境的检测和探知。

华为 Data Agent 智能体工具调用实践

华为 Data Agent 是基于华为 AI 应用引擎开发的数据智能体，采用多种类型工具，实现“数据驱动+业务洞察”的新模式。Data Agent 融合结构化和非结构化数据，基于开源和闭源大语言模型、MCP 工具、知识库、元数据、数据权限管控、数据查询分析、多智能体协作等核心能力，通过深度理解业务语境，主动思考拆解数据任务，自动调取各种工具处理企业数据、智能深度分析、自动执行任务，全天候帮助业务高效挖掘数据价值，支持业务用户通过自然语言等方式便捷完成面向行业应用的各种数据业务场景，如数据管理、数据分析、数据营销等。相比传统工具和普通问数智能体，Data Agent，不仅能回答“是什么”，更能推导“为什么”和“怎么做”，帮助用户制定可以落地执行的策略和解决方案。



来源：华为技术有限公司

图 3 华为 Data Agent 总体架构图

2. 任务执行确保动作及时响应

执行能力是任务落地的关键，也是智能体将决策转化为行动的重要阶段。执行包含任务调度、行为控制和结果评估三个核心部分。其

中，任务调度确定执行顺序与优先级；行为控制实现动态调整与反馈响应；结果评估提供优化反馈。常用执行策略包括行为树、模型预测控制和强化学习反馈机制等。**强化学习在行动决策中发挥核心作用。**Q-Learning、SARSA、DQN 等强化学习算法，让智能体在复杂或未知的环境中探索并利用，从而找到最优路径来达成目标。策略梯度算法通过直接优化策略函数提升决策质量。深度强化学习结合神经网络，使智能体能在复杂状态空间学习更高效策略。

（四）通信协议夯实原生基石，加固多体协同底座

随着智能体在感知理解、任务执行和自主决策等能力上的持续突破，单一智能体已难以满足复杂系统性任务的执行需求。多智能体系统正逐步成为构建智能生态的重要支撑形态，通过智能体通信协议、多智能体协同机制的规范化支撑，实现了高度自治、高适配性的创新协作模式。在此基础上，智能体之间不再是孤立的“工具节点”，而是具备感知、协同、进化能力的“智能单元”，可按需重构系统能力、按任务动态组网，推动人机协同、人机共创、人机共演的多维交互模式持续演进，重塑智能系统的基础交互范式与服务逻辑。

1. 通信协议开启智能协同新范式

通信协议通过提供标准化接口和通信机制，显著提升模型及智能体的协作能力，推动人工智能生态标准化与开放化。通过统一的通信协议和协同规则，实现异构智能体间的无缝对接，提升了系统的信息流通效率和任务处理能力。**模型与工具之间的通信方面**，通过统一标准化接口，解决模型与外部工具、数据的兼容性问题，实现“一次开

发，全生态通用”。MCP 采用“客户端—服务器”架构，实现对大模型的高效管理和外部资源调用，降低系统集成复杂度，拓展模型能力边界。智能体和智能体之间的通信方面，通过标准化接口打破智能体之间的孤岛，支持动态服务发现、任务分配、数据交换和实时协作，使得智能体之间的通信和协作变得更加高效、安全和标准化，为构建复杂的多智能体生态系统提供了坚实的基础。A2A 协议强化了智能体之间的点对点通信能力，构建起独立智能体间的通用消息收发机制，增强智能体间的自治通信与平台级联动能力。

2. 多智能体交互协同提高复杂任务处理能力

大模型为智能体提供了强大的自然语言理解和生成能力，也赋予其交互和协同能力，为多智能体系统带来了革命性突破。多智能体交互与协同涉及多种协作类型、协同策略和多样的通信结构。多智能体协作类型方面，可分为合作型、竞争型、混合型。在合作型协作中，智能体通过目标对齐实现协同增效。在竞争型协作中，竞争机制通过对抗性互动提升系统鲁棒性。而在实际的应用中，多层次混合型协作更为常见。多智能体协同策略方面，可分为规则驱动、角色驱动、模型驱动。在模型驱动策略下，智能体接收输入信息，根据环境状态的感知，共同协作目标，做出概率化的决策。这种方法赋予智能体系统高度灵活性和鲁棒性，也带来了更高的复杂性，设计和部署难度较大。多智能体通信架构方面，可概括为集中式拓扑、分布式拓扑、分层式拓扑三大类。集中式架构是一种所有智能体均与中央智能体相连的实现方式。分布式架构通过将控制权和决策权分散至各智能体。分层式

架构中采用层级化架构，各层级智能体具备特定功能，主要与同层或相邻层智能体交互。

3. 多智能体交互系统是实现高效协同的核心引擎

多智能体交互系统是支撑智能体协作任务高效完成的核心机制，具备分布性、自治性、协同性与适应性等关键特征，能够胜任复杂动态环境下的任务执行与反馈闭环。在任务协同与分工方面，多智能体系统能够将复杂目标自动拆解为多个子任务，并基于能力匹配将其分配给具备相应功能的子智能体处理，形成高效的并行任务执行网络。在智能协同决策机制方面，系统引入“规划—执行—验证”三代理协同模式。规划智能体负责任务解析与路径设计，执行智能体负责具体行动执行，验证智能体实时监控任务执行结果并进行偏差校验，保障任务流程的完整性与可控性。在任务动态调整能力方面，多智能体交互系统能持续监测与验证，当执行过程中遇到问题时，系统还能自动调整计划，保证任务继续推进，无需人工监控与干预，系统可自主完成从任务分解到成果交付的全链条操作。通过这一整套机制，多智能体交互系统实现了从任务输入、智能规划、分布执行、过程验证到结果交付的全流程闭环，显著提升了智能体在复杂场景中的协同效率与自治能力。

多智能体协同工作体系-故障监控专家

故障监控专家是由策略智能体、故障识别智能体、故障分析智能体、调度执行智能体、评估智能体、知识助手和数据助手构建的“5Agent+2Copilot”多智能体协同工作体系，与故障监控业务流程深度融合。

围绕星云通信大模型、多智能体协同构建故障监控专家，赋能网络故障事件端到端智能化管理，提升故障事件闭环效率，减少一线人工参与工作量，缩短故障处理时间。自研的自优化智能体协同双向去噪方法，将多智能体间协同准确率提升至 98%，直接使故障历时下降 15 分钟；通过故障因果思维链技术，重塑复杂故障分析能力，实现横向跨专业网故障定界定位准确率由 70%提升至 90%以上，纵向单专业网故障定界定位准确率提升至 92%以上。同时基于 PC 端、掌上端双触点，推出首个掌上大模型监控专家，实现一线运维人员与系统使用语言方式交互，降低其工作负荷 30%。



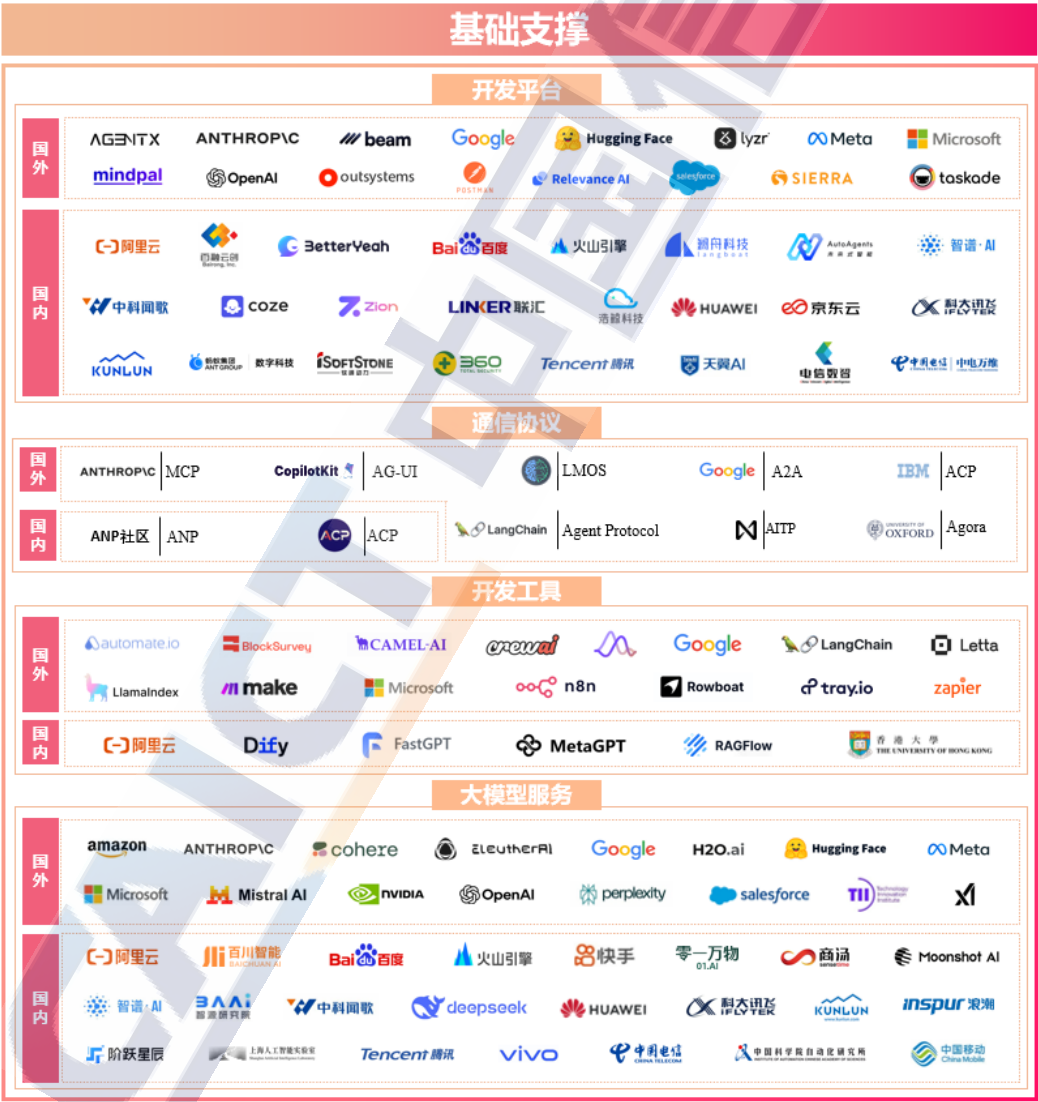
来源：中兴通讯股份有限公司

图 4 故障监控专家智能体

三、智能体产业应用

（一）平台工具加速落地进程，构建全域生态体系

随着智能体技术从理论迈向产业实践，平台化工具与生态协同成为规模化落地的核心驱动力。低代码框架、模块化组件和标准化接口构成的智能体平台降低了企业应用门槛。在垂直领域，智能体通过场景化功能模块拓展大模型应用边界，通过多智能体协作和开放生态平台打通数据孤岛，加速资源整合，推动产业从单点智能向全域协同升级。



来源：公开资料整理

图 5 智能体基础支撑图谱

1. 智能体降低了原生应用的开发与部署门槛

智能体技术的快速发展得益于底层开发框架与工具的成熟，大幅降低了开发门槛和部署成本。在开发和部署门槛方面，智能体框架通过提供预构建的组件、标准化的架构和易用的工具，简化了智能体的开发。低代码/无代码开发平台、自动化部署工具链和标准化接口协议的普及，简化了智能体流程，大幅降低了智能体原生应用的开发门槛。在分级部署方面，可根据业务需求灵活适配不同的量级，从轻量级应用到全域智能体平台均可高效落地。在轻量化部署场景中，可以基于 RAG 框架构建智能体，实现快速落地应用。在复杂业务场景中，可以进行全栈化部署，搭建全域智能体平台，整合千亿级模型与全域数据，实现从流量触达到转化提效的商业闭环。

蚂蚁智能体开发平台

Agentar 智能体开发平台依托其先进的大模型技术，为企业和开发者提供了一站式、多模型兼容的 AI Agent 开发解决方案。该平台以“低门槛、高扩展、全链路支持”为核心理念，覆盖从智能体创建、调试、部署到持续优化的完整生命周期，通过应用创建、知识管理、 workflow 编排、工具市场、智能体评测与开放 API 六大核心模块的深度整合，助力企业快速构建智能化业务场景并实现规模化落地。

Agentar 的核心优势在于其全链路能力的无缝衔接。用户无需编程基础，即可通过可视化界面，通过 Prompt 配置与行为定义，快速构建专属智能体，显著降低 AI 开发的技术门槛。在复杂业务流程的自动化方面，Agentar 的 workflow 编排功能突出。平台还提供了丰富的工具集与 MCP 服务，不仅包含数据处理、API 集成等基础工具，更支持自定义模块的自由开发与扩展，使智能体能深度适配个性化业务需求。



来源：蚂蚁区块链科技（上海）有限公司

图 6 蚂蚁智能体开发平台

2. 智能体平台工具丰富了大模型的应用体验

近年来，国内智能体开发平台凭借功能创新、场景适配和开源融合的特点，不断丰富大模型应用场景。功能创新上，平台具有多模态交互、提示词编排、知识库集成等创新性功能，能够降低智能体开发门槛，优化开发者使用体验，提高智能体研发效率。**场景适配上**，平台提升智能体场景使用体验，通过 API 调用可以轻松完成智能体在不同场景的集成和应用，为知识密集型场景深度赋能，为开发运维型场景降本增效。**开源融合上**，平台通过融合各类开源大模型、集成开源技术，提升智能体性能，加速平台技术进步，为智能体开发和落地应用奠定坚实基础。通过智能体开发平台创建各类智能体应用，不仅深化大模型在医疗、金融、教育、制造等领域的应用，还拓展在智慧城市、文化创意等新兴场景的应用。

“交心”智能体平台

交通银行深度践行国家人工智能发展战略，依托千卡异构算力集群、千亿级金融大模型算法矩阵和海量知识数据支撑，构建“1+1+N”人工智能建设框架体系并自主研发“交心”智能体平台。平台以“降成本、控风险、创价值、优体验”为导向，围绕统一 AI 云底座、统一 AI 训练、统一 AI 服务的建设目标，打造全栈技术可控的企业级智能体平台。平台以“模型即服务”模式支持快速接入商用/开源大模型，提供多层次、多能力融合服务。依托插件、工具及知识库组件，显著降低智能体构建门槛，助力全行业务、技术人员和应用系统快速构建 AI 应用。基于平台打造零售普惠、客服营运、风险授信、办公研发等领域端到端 AI 应用场景，目前已经落地近百个大小模型融合的应用场景，提升了行内金融业务的智慧化水平。



来源：交通银行股份有限公司

图 7 “交心”智能体平台架构

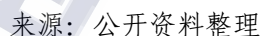
3. 智能体有助于构建全域生态体系

智能体多维能力的提升主要依赖通信协议标准化、动态协作框架

和生态体系建设三大关键机制。通过技术架构与协作规则的深度耦合，智能体生态正从“单点连接”向“全域联动”演进，推动行业从“数据孤岛”向“价值网络”转型。**通信协议标准化是基础**，应用方需根据跨平台互联需求设计统一通信协议，采用 MCP 或 A2A 等标准化接口方案，以解决异构系统间的兼容性问题。**动态协作框架是核心**，应用方需针对复杂场景设计多智能体协作机制，实现实时环境下的策略调整，包括任务分解与动态规划、角色扮演与协同推理等能力。**生态体系建设是保障**，通过开放标准与开源生态融合，推动产业链资源高效整合，最终形成跨行业、跨场景的全域生态体系，加速实现技术普惠与产业升级。

（二）通用场景实现百花齐放，助力实现降本增效

当今数字化浪潮的推动下，智能体技术正迎来前所未有的发展机遇，其通用场景的应用呈现出百花齐放的繁荣景象，应用价值在规模化实践中得到充分验证。不仅极大地提升了生产效率与服务质量，降低了人力成本与资源浪费，还为人们的生活带来了前所未有的便捷与舒适，推动着社会向更加智能化、高效化、人性化方向迈进，展现出强大的生命力与广阔的发展前景。



1. 智能体应用场景持续拓宽

1. 智能体应用场景持续拓宽

27

华为 Operator 智能体

华为 Operator 智能体是基于华为 AI 原生应用引擎开发的通用任务智能体，是一个能自动执行多步骤 workflows、智能调度资源并实时反馈任务状态的“AI 任务执行管家”。Operator 智能体可以基于用户输入的自然语言指令，基于大模型的视觉识别能力和通过强化学习实现的高级推理能力，以及内置的搜索、分析、生成和通过外部 MCP Server 扩展等工具，使得 Operator 能够模拟人类操作计算机的方式，包括浏览网页、使用鼠标和键盘进行交互等，并能自助规划和分解任务，从而自主的完成一系列复杂的在线任务，助力用户实现更加智能化的生活方式和工作流程。



来源：华为技术有限公司

图 9 华为 Operator 智能体总体架构图

2. 智能体应用价值逐步明晰

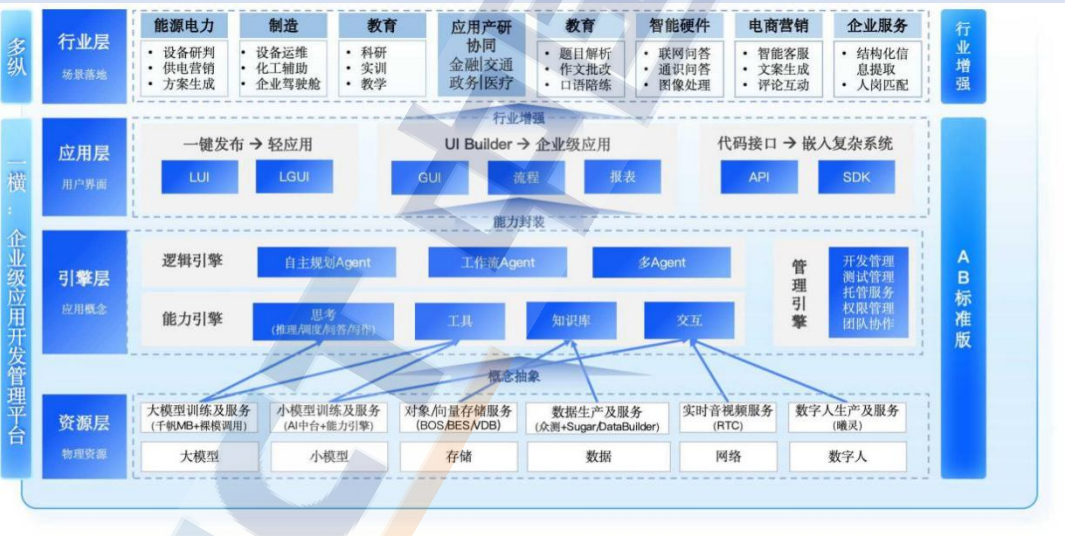
智能体打通业务场景和人类交互的桥梁，正在承接日益复杂的提质增效需求。智能客服方面，智能体能够自动识别和理解用户的问题，并提供准确、及时的回答，有效缓解人工客服的压力，降低人力成本，提升客户满意度。同时还能通过分析用户咨询数据，挖掘潜在需求，为企业提供有价值的市场信息。智能运维方面，智能体可以通过对系

统日志、性能指标、网络流量等多维度数据的实时监测和分析，快速发现潜在的故障隐患，实现故障的快速定位和修复，减少系统停机时间，保障业务的连续性，提高运维效率和质量。**风险管理方面**，通过对海量客户数据的深度分析，智能体能够构建精准的风险评估模型，能够提高风险识别的准确性和及时性，同时，基于数据驱动的风险管理模型能够不断优化和调整，适应市场变化和新的风险特征，增强企业的风险抵御能力，保障业务的稳健发展。**智能营销方面**，智能体可以动态调研市场数据，了解客户的产品需求，根据客户的偏好和行为提供个性化营销服务，进而提升营销效率、改善客户体验。

AppBuilder 智能体平台通用场景落地实践

千帆 AppBuilder 智能体平台作为企业级大模型应用开发平台，已在智能客服、供电营销、设备运维、方案生成等场景中应用，加速企业级大模型应用落地，提效大模型应用落地最后一公里。

在营销服务场景中，平台通过统一数智化底座，将 AI 能力带入“研产供销运、买卖用修服”数字化全场景；在智能助手场景中，智能体平台融合先进大模型技术和现有校园业务，构建便捷师生的校园助手；在智能客服场景中，落地云智一体全场景，孵化货币交易机器人、对话生成等 AI 原生应用；在办公服务场景中，平台内置 RAG 框架和专有知识库，实现对内构建办公应用、对外构建客服应用，辅助企业提高办公效率和服务质量。



来源：北京百度网讯科技有限公司

图 10 千帆 AppBuilder 智能体平台

（三）专用场景持续走深向实，加速推进数转智改

在当今数字化与智能化加速发展的时代，智能体专用场景的持续走深向实，正成为推动社会生产变革的重要力量。根据特定行业的需求和特点，智能体可提供高度个性化的服务，智能体通过深度融入各

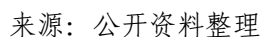


图 11 智能体专用场景应用图谱

1. 智能体向专用场景应用延伸

智能体专用场景的应用具备专业性强、定制化需求高的特点，被广泛应用到社会发展的各个领域。在金融领域，基于行业知识图谱构建决策智能体，可以提供智能投顾、智能信贷、实时反欺诈与风控、智能客服等服务，从而降低金融服务门槛，重塑金融业务流程。在医疗领域，智能体通过分析患者的病历数据、医学影像数据和实验室检测数据，帮助医生更快速、更准确地进行诊断和制定治疗方案。在政务领域，智能体能够自动处理重复性高、规则性强的政务任务，如材料审核、工单分派等，显著提升效率通过多智能体协同，打破部门间的数据壁垒，实现跨部门的高效协作。在教育与文化领域，相关智能体应用可以识别错题、自动生成同类知识点题目、定制学习路线、生

成教案等，提高老师和学生的工作学习效率，也可以通过 AI 数字人提供远程教学，文化创作等功能，赋能教育公平实践和文化遗产创新。

华为 Deep Research 智能体专用场景落地实践

华为 Deep Research 智能体是基于华为 AI 原生应用引擎开发的研究报告智能体，是一个“AI 研究报告专家”。Deep Research 智能体面向需要在金融、科学、教育等各个行业需要开展专业研究的分析人员或者某一方面需要进行深入分析的个人研究者。该智能体基于用户输入和研究主题，对研究课题进行拆解，自动浏览网页，利用推理来搜索、解释和分析互联网上相关的文档、图像等资料，最终生成一份专业的研究报告，并附有相应引用来源，便于用户验证和参考，显著提升用户研究分析效率。



来源：华为技术有限公司

图 12 华为 Deep Research 总体架构图

2. 智能体持续赋能企业智能化升级

智能体的行业应用促进企业运营效率提升，推动企业数字化转型迈向新高度。在工业制造领域，智能体通过边缘计算、智能协同分工等技术实现生产中智能分拣与物料管理、设备预测性维护、声纹监测、视觉质检、动态调度等，提升工厂生产效率和资源利用率。在城市管

理与公共服务领域，通过全域感知互联和多智能体分工协作，实现交通管理、社区治理、应急安全调度、环保监测、民生政务等公共服务的智能化运行，构建宜居、智慧的未来城市。**在汽车驾乘领域**，通过集成多模态感知与交互技术的智能体打造智能座舱，支持自动驾驶、车路协同、自然语言控制、预测性维护等功能，为用户带来个性化、情感化的智能驾驶体验。**在能源安全领域**，智能体应用通过智能勘探、设备智能监测、矿山智能风险研判、智能管网监测、智能调度等方式赋能能源生产、储运、消费等环节，推动能源行业的生产效率提升与安全范式创新。

木卫四蝴蝶大模型智能体汽车行业应用实践

木卫四依托“蝴蝶大模型”，深耕汽车垂直领域，构建了具备全链路能力的智能体服务平台。平台基于高质量汽车数据的持续积累与训练，融合数十万条安全因果链和数百万真实用车场景，具备对网络安全、数据安全、功能安全等多维异常的深度理解与推理能力。依托蝴蝶大模型的能力，已构建多类智能体服务，包括：VSOC运营智能体、威胁情报智能体、出行助手智能体、维修助手智能体等。

该平台以“蝴蝶大模型”为核心，融合代码生成、推理分析、知识问答、意图理解等能力，配套完善的法规知识库、垂类知识库及安全因果链数据，构建面向车辆全生命周期的智能服务体系。目前已覆盖 OEM、Tier1&2、车主、4S 店等关键场景。



来源：木卫四（北京）科技有限公司

图 13 蝴蝶大模型智能体服务平台

四、智能体问题挑战

（一）技术底座能力发展放缓，认知规划能力不足

1. 环境复杂动态多变，感知学习能力有待提升

当前智能体技术底座在复杂动态环境下面临五大核心挑战。一是多模态感知融合效率不足，视觉、雷达等多源数据实时协同处理能力弱，动态因素（如光照变化、突发障碍物）易引发融合延迟，跨模态信息冗余与冲突问题突出。二是实时学习框架滞后，传统深度学习难以支撑低延迟在线增量学习，突发场景下知识库更新与策略优化时效性不足。三是小样本与无监督学习薄弱，依赖监督学习的模型难以应对数据标注成本高、样本稀缺问题，工业场景中设备状态自主感知能力受限。四是动态环境仿真覆盖不足，现有虚拟测试平台（如自动驾驶）无法完全复现极端天气、人车交互随机性等真实复杂特性，模型泛化能力弱。五是感知-决策闭环协同性差，模块割裂导致数据噪声干扰决策，实时反馈与优化机制亟待突破。通过优化多模态融合算法，创新实时学习框架，强化小样本与自监督技术，构建高保真仿真环境，并深化感知-决策闭环协同，为智能体规模化应用奠定基础。

2. 幻觉模糊性风险潜伏，规划决策能力尚不成熟

基于大模型的智能体虽然取得了不错的成绩，但在推理、规划、创造性思维和决策等方面仍存在挑战。数据质量方面，智能体的运行依赖多维度数据支持，包括模型预训练数据、精调优化数据、推理调优数据、评测数据以及外部专业知识库与实时数据。然而，数据质量缺陷会直接影响智能体输出的准确性与专业性。推理和规划能力方面，智能体需具备基于目标的推理和规划能力，以高效处理复杂任务。由于智能体发展尚不成熟，技术能力具有一定的局限性，存在无法准确

预测未来的事件或结果、无法获取完整信息的可能性，进而导致智能体规划失效或产生不可预测的结果，面临不确定性和模糊性的挑战。**可靠性和透明可解释性方面**，当前智能体的“大脑”是基于大模型，可能出现“幻觉”问题。因此，需要结合 RAG 等一系列技术措施确保最终输出结果的可靠及可信。此外，在智能体落地应用中，不仅需要提升智能体运行过程的可观测性，以便监控其行为，更要增强决策过程的可解释性。

3. 协同交互尚未充分显现，竞争合作亟需平衡发展

在多智能体系统中，智能体之间的协作和竞争平衡成为了关键挑战。随着智能体数量的增加，如何处理智能体之间复杂的交互和协调，如何解决智能体之间的冲突和矛盾，如何评估智能体的表现和进步等问题亟待解决，以实现智能体之间的有效沟通和协作。**交互和协调方面**，智能体通过通信和信息共享来协调行动，共同完成任务。对智能体的协同性、一致性以及智能体之间的信任程度提出了高度的要求，要求智能体能够理解和遵循共同的规则和目标，并信任其他智能体将按照约定行动，以实现整体目标。**竞争方面**，每个智能体具备特定的目标、行为、信念和偏好，同时也受到环境的影响和约束。在实现共同的目标或任务的过程中，智能体之间可能存在利益冲突，过度竞争将导致资源浪费和效率降低，甚至可能引发系统的不稳定。**协作和竞争机制方面**，智能体需要在协作和竞争之间找到一个平衡点，以实现整体系统的最优效果。如何设计和实现多个智能体之间的协作和竞争机制成为多智能体系统的核心问题之一。**通信兼容性方面**，智能体可

能面临异构性导致的协议与标准碎片化问题，智能体应用过程中的信息孤岛和通信兼容性问题亟待解决。

（二）应用场景创新面临挑战，个性需求挖掘不深

1. 应用服务初显范式，纵深发展任重道远

不同行业对智能体的依赖程度有所不同，拓展智能体领域应用的深度和广度是一个具有挑战性的任务。应用广度上，基于大模型的智能体当前在多个下游领域中展现出强大潜力，但在创意产业、艺术等需要创造性思维和人类情感的领域应用相对较少。智能体需要进一步打破大模型到场景化落地应用的困局，应用广度有待进一步拓展。应用深度上，通过大语言模型的语言理解、知识推理、数据沉淀等能力和新型人机协同模式的赋能，为提高下游任务的执行力带来了更多的可能性。例如，在制造业、金融业、医疗保健等领域，可以通过自动化和智能化技术提高效率和质量，对智能体具有较高的依赖性。随着各环节的技术持续革新，智能体在不同垂直行业领域的应用进程有待加速，以提高人工智能应用生态的成熟度。

2. 交互需求愈发清晰，应用生态亟需完善

智能体根据用户的指令确定目标，通过与环境交互优化性能指标，要求智能体能够理解和执行复杂的用户指令。交互方式上，当前，大、小模型在应用层面进行交互的典型方式为插件模式，即将模型构建的应用封装为插件服务供其他模型调用。而大部分应用插件可通过智能体实现协同调用，以执行多模态场景任务，满足用户的需求。交互能力上，智能体不仅应探索数据安全性与个性化的平衡，还要提高大模

型的准确性、强化泛化性，进而深化智能体的交互关系形成智能体网络，丰富整体服务能力，拓宽服务应用广度，达到完善应用生态的目标。

3. 个性化需求鞭策，引领服务模式创新升级

随着智能体的广泛应用，个性化创新需求日益凸显。智能体的灵活性和适应性亟需提升，以满足用户个性化需求，带来更好的服务和体验。**灵活性上**，外界环境不确定性及动态性变化，促使智能体应具备高度的适应性和灵活性，拥有与时俱进的能力。部署具有专业知识和独特技能的智能体，使其在社会中扮演不同的角色，成为行业界的重点关注目标。**个性化上**，人工智能和智能体技术的不断发展和应用，促使用户对智能体服务或产品的期望不断提高，对满足特定用户、特定场景和特定任务的个性化需求日益增长。智能体需具备适应不同任务和场景的能力，通过个性化创新来满足用户个人偏好、习惯和需求，进而开发出更加智能化、个性化和实用化的智能体应用，为用户带来更好的体验和服务。

4. 复合型人才短缺，制约智能体应用发展进程

目前行业中缺乏既懂业务又懂 AI 技术的专业人员来有效地开发、部署和监管智能体，复合型人才面临着较大缺口。在智能体开发和定制阶段，需要相应人才能够理解领域的特定需求，并将需求转化为可执行的 AI 模型和智能体功能。**在智能体部署阶段**，将开发完成的智能体系统集成到现有基础设施的过程中，应确保其安全性、稳定性以及与其他系统的兼容性，需要相应人才有专业的 IT 和 AI 技能，并理

解行业的安全标准和监管要求。在智能体运营阶段，相关人员应能够监控智能体的运行状态，及时发现和解决潜在的风险和错误，评估其性能并根据业务变化进行调整和优化。

（三）安全治理体系仍不完善，标准规范构建不全

1. 隐私安全面临严峻挑战

随着智能体的普及，如何防止恶意攻击、保障用户隐私安全成为亟待解决的问题。安全性和隐私性是智能体的关键特性，对其稳定运行和对用户及社会的保护至关重要。数据隐私方面，智能体除了具备能力强大的模型底座支撑，还需要专业化数据的加持，以提供个性化的服务和建议。智能体需要访问和利用用户的个人信息，由此衍生出个人数据、隐私相关的问题，引发业界对数据隐私和安全性的担忧，在智能体应用过程中，企业自身数据资源也可能因为技术漏洞、协作机制设计缺陷或管理流程不足等原因面临暴露风险，亟需通过技术与管理结合的方式系统性解决。安全性方面，智能体落地应用面临着多种网络安全风险和攻击手段，如提示词注入、数据投毒、智能体权限滥用、智能体中间人攻击等。制定合理的安全保障机制，消除漏洞、遭受攻击或数据泄露等安全隐患，增强智能体的信任度和控制力，为个人信息安全保驾护航，避免对用户或社会造成损害。

2. 伦理法律迎来双重考量

智能体的应用也面临着伦理和法律方面的挑战。伦理方面，伦理和责任是智能体可信应用的重要原则，当智能体在决策过程中需要权衡不同的利益时，如何确定其决策是否公平、公正、透明、可靠成为

重要的议题。智能体的决策过程需要符合伦理标准，防止算法偏见，确保所有人群都能公平享受 AI 带来的便利，避免因种族、性别、年龄等因素产生偏见或不公正的待遇。**法律方面**，责任归属是智能体的关键议题，人与智能体协同中的责任归属不清晰或不公正也会带来严重后果，引发法律责任问题。例如，在自动驾驶汽车发生事故时的责任承担问题；在真人形象模仿应用中，如何区分真人和以他为原型制作的智能体权利界限，可能涉及人格权及名誉权的纠纷；在研发作为生产力工具的智能体过程中，可能涉及很多包含专利、著作权、商业秘密的材料，引发知识产权相关的问题，需要在立法和监管方面加以完善。

3. 社会环境存在多维掣肘

智能体的兴起深刻影响社会的多个层面，可能带来就业市场的重塑和技术失控的风险。智能体的普及将有可能进一步加剧技术性失业。智能体通过利用个性化和专业化的数据进行训练，打破了通用大模型在情绪价值和专业知识要求较高的场景上限制，业务表现得到大幅跃升，智能体应用更加广泛，将带来技术性失业的问题，可能会导致一些传统岗位的消失，引发就业结构的深刻变化。面对这一挑战，社会应未雨绸缪，通过教育体系改革、职业培训转型等方式，帮助劳动力适应新经济环境，促进人与机器的和谐共存。**智能体的广泛应用可能会导致各种不可测的失控风险。**如何确保智能体的发展始终处于人类的可控制范围内，防止技术滥用或误用带来的灾难性后果，是科研人员、政策制定者和全社会必须共同克服的难题。

五、智能体发展建议

（一）全方位加强大模型攻关，夯实智能技术底座

智能体技术的发展高度依赖大模型的突破与支撑。随着多模态交互需求的爆发式增长与行业数字化转型的纵深推进，智能体迈向更高水平须全方位加强大模型攻关，围绕“自主决策能力提升-跨领域学习能力增强-高性能设施支撑”三大维度构建发展新范式。

1. 自主性迈向高级别，情感理解和表达能力更佳

智能体的自主性提升是技术演进的重要方向，情感交互能力是智能体迈向高阶智能的关键标志。自主性方面，需要着力突破其在复杂场景下的独立决策与动态适应能力，通过强化大模型的认知推理与意图理解能力，推动智能体从被动响应向主动规划转变，实现任务分解、资源调度、动态优化的全流程自主闭环。在此过程中，需要深度融合多模态感知与语义理解技术，构建基于上下文的学习机制，使智能体能够精准识别用户意图并动态调整行为策略。情感交互方面，需要重点攻关情感计算与生成技术，通过语义情感分析、非结构化数据建模、多模态情绪识别等手段，增强智能体对人类情感状态的感知深度。同时，需要优化生成模型的共情表达能力，在自然语言交互中融入情感逻辑与语境适配能力，实现更具人性化的对话体验。

2. 跨领域学习能力更强，塑造智能体发展新格局

智能体的跨领域泛化能力决定其应用边界与生态价值。技术路径方面，需要以大模型为核心底座，突破领域知识迁移与异构数据融合的技术瓶颈，构建统一的知识表征框架与自适应学习范式。强化跨模

态信息协同与知识蒸馏能力，打通文本、语音、图像等多源数据的语义关联，形成全域知识图谱支撑下的推理链路。**场景适配方面**，通过多任务联合训练、元学习、增量学习等机制，增强智能体对多样化场景的快速适配能力，实现从单一功能向综合服务能力的跃迁。设计轻量化、模块化的模型架构，支持智能体在边缘计算与云端协同场景下的灵活部署，推动智能体从垂直领域向跨行业融合应用延伸。**跨领域学习方面**，智能体应具备更强的跨领域学习能力，实现快速适应新的环境和任务，降低从头开始学习的复杂性。改进算法、增强学习机制、开发更复杂的记忆系统等多种方式的积极探索，以提升智能体的能力价值。

3.基础设施需求驱动，加速数字化升级进程

智能体的规模化应用需要新型数字基础设施作为基础保障。为满足大模型训练与推理需求，需围绕算力、数据、网络三大核心要素构建高性能技术底座。**算力层面**，注重架构创新与技术突破，以“算网融合”为目标，助力超级算力系统发展。应重点突破异构计算资源调度与能效优化技术，推动智算中心向集约化、绿色化方向演进。**数据层面**，完善多源数据治理体系，强化数据清洗、标注、安全防护的全流程能力，释放高质量数据要素价值。**网络层面**，探索低时延、高可靠的边缘云协同架构，满足智能体实时交互与分布式部署需求。同时，应需加速人工智能与 5G、物联网、区块链等新一代信息技术的融合创新，打造“云-边-端”一体化的智能基础设施体系，并通过标准化接口与开放平台建设等，降低智能体开发与集成的技术门槛。

（二）多领域促进智能体落地，推动应用普惠向实

智能体正从技术验证迈向规模化落地阶段，其价值释放的关键在于突破垂直领域壁垒、激活场景创新潜能、构建普惠服务生态。通过跨学科知识交叉重构技术范式，依托新型场景激发服务创新活力，聚焦社会核心诉求释放技术普惠价值，推动智能体技术与产业需求深度融合，形成“技术赋能-场景驱动-社会反哺”的良性循环。

1. 应用领域纵深拓展，促进跨学科交叉融合

智能体的深度渗透将突破传统行业边界，促进跨学科协同创新。基础理论方面，构建“AI+X”的跨界范式，将人工智能与哲学、社会学、伦理学等人文领域深度结合，构建“理心合一”的理论框架，重点攻关认知科学与社会学交叉的社会模拟与群体智能算法。知识融合方面，从中国传统文化中提炼哲学智慧，将其转化为通用人工智能的认知逻辑与伦理准则，将人文思想嵌入智能体的决策链路，使其具备类人价值观与社会协作能力，突破传统 AI 工具性局限。协同机制方面，建立跨学科动态学习体系，推动智能体与社会科学实验深度结合，通过模拟与建模，让文明、社会、经济与政策等可以进入可验证的科学范畴。探索人机协同的新型科研范式，将智能体的数据分析能力与人类专家的经验深度结合，攻克跨领域系统难题。

2. 探索新型应用场景，激发服务创造新生机

开拓智能体的新型应用场景是激发创新活力的重要途径。场景创新方面，突破传统服务边界，重点聚焦元宇宙、数字孪生、具身智能、群体智能协同等新兴领域，探索智能体在虚拟空间构建、物理世界映

射、人机环境交互中的创新应用模式。**行业应用方面**，建立领域知识图谱与动态学习机制，实现专业领域规则与通用智能能力的有机耦合，强化智能体在医疗诊断、气候预测等专业场景的精准服务能力。**社会实验方面**，通过大型社会模拟器、通用智能体等平台模拟经济社会的运行规律，探索人口政策、文明演化等复杂问题的可验证模型，将文科研究从定性分析转向定量实验。**落地实施层面**，在智慧城市、智能制造等领域开展小范围智能体应用实验，加速场景原型验证与规模化复制，形成“技术建模-社会验证-决策优化”的闭环。

3.挖掘应用创新空间，推动社会繁荣进步

智能体的应用目标是服务社会发展。民生领域，需要以公共利益为导向，以智能体应用为驱动引擎，进一步推动公共服务均等化、医疗资源普惠化、教育机会公平化、文化传承数字化等，实现资源高效触达与权益公平覆盖。**社会治理领域**，探索智能体驱动的新型社会治理模式，实现风险预警、政策模拟、应急响应的智能化升级，提升社会运行效能与韧性。智能体将凭借自身能力特性，在解决环境问题、提高资源利用率、促进公平正义等方面发挥关键作用，持续推动社会进步。**可持续发展领域**，深化绿色低碳与循环经济场景赋能，推动技术红利向生态效益转化，构建人与自然和谐共生的智能技术生态，为全球可持续发展提供创新范式。

（三）高标准引导可信赖发展，对齐人类价值偏好

智能体技术的应用引发了对安全、伦理和法律的深刻思考。为确保其发展始终服务于人类福祉，需要从安全可信、伦理道德和法律政

策三个维度构建全方位的治理体系。首先，通过技术与机制的协同创新，确保其在复杂环境下的稳定性和可靠性。其次，以伦理道德为约束，推动智能体与人类价值观深度对齐。最后，通过健全法律政策体系，为智能体技术的发展提供明确规范和引导。

1.安全可信多措并举，为智能体发展保驾护航

构建智能体安全可信保障体系是技术落地的基础。技术研发方面，加强智能体的安全性测试与验证，建立贯穿研发、训练、推理和应用全过程的测试框架，确保智能体在复杂环境下的稳定性和可靠性。同时，通过攻防测试等手段，识别潜在风险并及时优化。监测预警方面，构建智能体风险监测与防控体系，建立风险预警和应急响应机制，确保在突发情况下能够快速采取措施，降低损失。通过技术、标准和监测的协同发力，筑牢智能体发展的安全基座。隐私安全保护方面，重点关注隐私保护和信息安全，保障过程可控和结果可信。相关机构和企业应加强隐私和安全保护，重点关注数据隐私、安全性、鲁棒性等问题，充分保障用户的数据和隐私安全。智能体在收集、处理和使用数据时必须遵守相关法律和道德准则，并采取有效措施确保数据的安全性和隐私性。标准制定方面，制定涵盖功能安全、数据安全的多维度标准体系，明确技术要求和规范，建立智能体技术产品认证制度，引导行业在开发和应用中遵循统一的安全准则。人机协同共生方面，未来，智能体将更加注重与人类的和谐共存，发展人机交互界面，提升智能体的可解释性和透明度，促进人类对 AI 的信任与接受度。

2.伦理道德标准约束，实现人类价值观对齐

建立智能体伦理道德标准体系是技术向善的关键。理论研究层面，研发支持人类价值偏好学习的算法框架，建立人机价值协商机制，界定其在社会、文化、经济等领域的适用范围；同时，深入探讨如何通过技术手段保障智能体的价值观公平性，避免因数据偏见或算法缺陷导致的价值冲突。**技术创新层面**，采用人类反馈强化学习、可扩展监督学习等技术方案，突破智能体决策过程可视化技术，构建人机信任协作模式，确保智能体的训练目标与人类价值目标保持一致。**应用实践层面**，制定智能体技术应用伦理指南，建立伦理审查委员会制度，在智能体落地应用前进行严格的价值对齐评测，通过安全测评、可解释性分析和人类价值验证等方法，充分识别潜在风险。通过理论研究、技术创新、伦理审查等协同发力，确保智能体技术发展符合人类根本利益与伦理规范。**社会道德层面**，为应对伦理道德挑战，技术人员应考虑实施行为和道德约束，遵循人类人工智能设计的最佳实践。基于智能体的系统设计和应用应考虑到不同群体的需求和利益，避免歧视和不平等的现象。制定伦理准则时需要确保智能体的行为公平、包容并尊重多样性。

3.健全法律政策体系，为人类福祉做出贡献

完善智能体技术法律政策体系是规范发展的保障。法律法规方面，需要加快制定针对智能体的专项法规，明确技术应用的法律边界，完善数据产权、算法责任等法律制度，规范其在数据隐私、知识产权、算法透明性、致害责任认定等方面的行为。**监管机制方面**，为确保智能体的健康发展，政府和相关机构应就相关问题展开充分的研究和讨

论，制定智能体发展的政策引导框架，平衡技术创新与社会风险，形成政府监管与行业自律协同治理框架，确保智能体技术的开发和应用符合社会公共利益。**国际合作方面**，应积极参与全球智能体治理框架的构建，推动国际社会在智能体技术标准、伦理规范和法律政策上的协调与共识，共同应对智能体技术带来的全球性挑战。通过法治建设、制度创新与全球合作，引导智能体技术向促进公平、提升福祉方向发展，确保其始终服务于人类共同福祉。

缩略语

AI	Artificial Intelligence	人工智能
ANP	Agent Network Protocol	智能体网络协议
API	Application Programming Interface	应用程序编程接口
A2A	Agent-to-Agent	智能体对智能体
CoH	Chain of Hindsight	回顾链
COT	Chain of Thought	思维链
DQN	Deep Q-Network	深度 Q 网络
IT	Information Technology	信息技术
IoT	Internet of Things	物联网
LLM	Large Language Model	大语言模型
LLM+P	Large Language Model plus Planning	大语言模型与传统规划方法相结合
MCP	Model Context Protocol	模型上下文协议
MQTT	Message Queuing Telemetry Transport	消息队列遥测传输
RAG	Retrieval Augmented Generation	检索增强生成
SQL	Structured Query Language	结构化查询语言
ToT	Tree of Thoughts	思维树

编制说明

本研究报告自 2025 年 3 月启动编制，分为前期研究、框架设计、文稿起草、征求意见和修改完善五个阶段，面向智能体落地应用的技术供应方、服务集成方、应用需求方开展了问卷调查和深度访谈等工作。

本报告由中国信息通信研究院人工智能研究所联合华为技术有限公司牵头撰写，撰写过程得到了人工智能关键技术和应用评测工业和信息化部重点实验室的大力支持。参编单位包括南京新一代人工智能研究院有限公司、中国邮政储蓄银行股份有限公司、交通银行股份有限公司、北京百度网讯科技有限公司、蚂蚁区块链科技（上海）有限公司、北京京东世纪贸易有限公司、北京快手科技有限公司、联通（广东）产业互联网有限公司、中移（杭州）信息技术有限公司、北京大学武汉人工智能研究院、河北工业大学、煤炭科学研究总院有限公司、木卫四（北京）科技有限公司、中国科学院文献情报中心、中兴通讯股份有限公司、北京数势云创科技有限公司、亚信科技（中国）有限公司、北京至顶信息技术有限公司。

中国信息通信研究院 人工智能研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62301618

传真：010-62301618

网址：www.caict.ac.cn

