

数据智能体研究报告

(2026 年)

(征求意见稿)

中国互联网协会人工智能工作委员会
中国信息通信研究院工业互联网与物联网研究所

2026年8月

版权声明

本研究报告版权属于中国互联网协会人工智能工作委员会，并受法律保护。转载、摘编或利用其它方式使用本白皮书文字或者观点的，应注明“来源：中国互联网协会人工智能工作委员会”。违反上述声明者，本单位将追究其相关法律责任。

编制说明

编制单位：（排名不分先后）

中国互联网协会人工智能工作委员会、中国信息通信研究院、杭州数美科技有限公司、鹏城实验室、北京邮电大学、中国电子技术标准化研究院、蚂蚁科技集团股份有限公司、安途金信商业保理有限公司、中交武汉智行国际工程咨询有限公司、中国移动通信集团有限公司研究院、朗坤智慧科技股份有限公司、中移（苏州）软件技术有限公司、中电信数智科技有限公司、上海数据发展科技有限责任公司、湖北省电子信息产品质量监督检验院、杭州网易智企科技有限公司、天津大学、网宿科技、福州市晋安区琴国生态科技有限公司、数川大数据（云南）有限公司、北京奥思工联科技有限公司、江西仟得文化传播有限公司、江西时刻互动科技股份有限公司、政达研学（雄安）科技有限公司、北京万维盈创科技发展有限公司、中国航空结算有限责任公司、可达智灵（北京）科技有限公司、数易元（杭州）数字科技有限公司、沈阳创链信息技术有限公司

编写组主要成员：（排名不分先后）

施建锋、刘远超、尹子航、刘阳、尹铃元、王鲲、唐昊然、蒋俊、何帅、杨玉虎、唐波、李盈锦、王智信、陈坦坦、冯煜、张金国、赵璐、潘惠梅、闫超众、胡进宝、孟长庚、徐煦、邹巍、衣丽萍、殷晓、苏蕾权、罗岗、李悦晴、熊苏清、王鑫、张宇、孙驰野、王行远、胡杰英、李春艺、张森、周兴、尹晓刚、付强、张凡、王彬、唐泉、唐续豪、高峰、庞川、刘永恒、王波、李铮、王隼、王一帆、赵有金、

宋博文、彭冬

中国互联网协会人工智能工作委员会

编写组联系方式：

施建锋（中国信息通信研究院）

18811593392

shijianfeng@caict.ac.cn

目录

一、 数据智能体的内涵定义	1
(一) 数据智能体发展环境	1
(二) 数据智能体主要解决问题	2
(三) 数据智能体定义	3
(四) 数据智能体核心能力特征	4
(五) 数据智能体带来的产业核心变革	6
二、 数据智能体的形态及演化	7
(一) 数据智能体与其他数据工具的差异	7
(二) 数据智能体主要形态	8
(三) 数据智能体能力演化	13
三、 数据智能体的技术体系	17
(一) 数据智能体的参考架构	17
(二) 数据智能体的关键技术体系	21
四、 数据智能体的产业生态	35
(一) 产业生态全景图谱	35
(二) 商业模式探索	37
五、 数据智能体的场景应用	39
(一) 数据智能供给与治理	39
(二) 数据智能加工与交付	41
(三) 数据智能分析与运维	42

(四) 数据智能安全与合规.....	43
六、 数据智能体的挑战及展望.....	45
(一) 产业落地核心挑战.....	45
(二) 技术发展趋势.....	46
(三) 行业生态未来前景展望.....	47

中国互联网协会人工智能工作委员会

一、数据智能体的内涵定义

（一）数据智能体发展环境

数据成为继土地、劳动力、资本、技术之后的第五大生产要素，当前数据要素的乘数效应正通过智能化手段加速释放，成为培育新质生产力的核心放大器。我国数字经济核心产业增加值占 GDP 比重已突破 10.5%；2025 年全国数据生产总量突破 50 泽字节（ZB），数据资源规模跃居全球首位，累计建成高质量数据集超 10 万个、总规模达 890 拍字节（PB）。其中，人工智能训练和推理所用的数据总量达 199.48 艾字节（EB），较上年增长 42.86%，**推理数据量以 101.34EB 的规模首次超过训练数据量**，人工智能对高质量数据的需求呈爆发式增长，但传统数据要素利用体系仍面临**供给不足、流通不畅、应用不深、安全合规薄弱**等问题，导致海量数据资源难以转化为现实生产力，制约了人工智能产业的规模化落地与深度赋能。

与此同时，人工智能产业正迎来从“大模型能力竞赛”向“智能体落地竞赛”的范式转移，数据智能体作为 AI 技术在数据要素领域的核心落地方向，成为产业竞争的新焦点。2026 年，国务院印发《关于推进服务业扩能提质的意见》，明确支持采购大模型与智能体服务，打通了智能体从技术研发到产业落地的关键链路；工业和信息化部、国家数据局联合启动 2026 年“模数共振”行动，提出要“打造能够符合场景特殊应用需要的专用模型或特色智能体”；国家网信办、国家发展改革委、工业和信息化部联合印发《智能体规范应用与创新发展实施意见》提出要“强化应用牵引”，列举了智能体的典型应用场景。国家

数据局印发《关于推进行业高质量数据集建设行动的实施方案》，提出打造一批数据赋能智能体解决实际问题的典型案例。



图 1-1 数据智能体的定位

数据智能体是贯通“数据资源-数据基础设施-智能体-应用场景”全链路的**核心枢纽节点**。向下，它是新一代数据基础设施的核心智能组件与数据要素**价值转化引擎**，激活全域沉淀数据的生产要素属性；向上，它是通用及垂直行业智能体的通用**底层能力基座**，为各类智能体提供标准化、高质量的数据能力支撑；向外，它是全行业数字化转型场景的**核心执行载体**，将数据能力转化为可落地的业务价值，本质上更是系统性破解数据要素全链路核心痛点的**核心技术抓手**。依托全链路自主能力，数据智能体推动数据处理范式从人工规则驱动向智能自主驱动根本性跃迁，进入产业落地关键窗口期，正在重构全球数据技术体系与产业组织模式，为数字经济高质量发展提供坚实底层支撑。

（二）数据智能体主要解决问题

传统数据要素利用体系长期依赖人工规制与静态配置，在数据供

给、流通、应用、治理全链条存在深层次瓶颈，难以适配数据规模化增长与业务实时化运行需求，核心痛点集中体现为四方面：

一是**数据有效供给不足**：多源异构数据的采集、加工与标注过程过度依赖人工干预，处理周期冗长、场景适配性弱、资产转化效率低，传统人工模式根本无法支撑海量数据的实时化、规模化加工需求，大量原始数据难以高效转化为标准化可用资产，无法形成稳定高效的数据供给能力。

二是**数据要素流通不畅**：数据呈现碎片化孤岛化分布，跨部门、跨主体流通高度依赖人工对接与静态配置，价值难以评估、检索匹配效率偏低、流通链路梗塞频发，人工对接模式无法打破跨域数据壁垒、实现供需精准匹配，数据要素难以实现跨域顺畅流通与高效协同。

三是**数据价值挖掘与应用效能偏低**：数据分析与业务支撑依赖于预设规则与静态分析，响应滞后、价值挖掘浅表化，传统静态分析模式无法支撑数据自主分析、高效科研创新与实时运营决策，无法将数据价值有效转化为决策支撑与核心竞争力。

四是**数据安全和合规治理能力不足**：数据质量管控、合规审核与安全防护主要依赖事后人工核查和被动处置，动态风险防控能力薄弱，隐私泄露、数据质量失真、合规越权等风险频发，难以实现数据全生命周期安全可信、合规可控。

（三）数据智能体定义

数据智能体（Data Agent）是 AI Agent 在数据领域的专业化应用形态，以数据及其相关资源为作用对象，以大语言模型等基础模型为认知与决策核心，融合多模态语义理解、知识推理、工具调用与任务

编排等能力，能够根据任务目标自主规划、执行和优化数据采集、标注、治理、发现、传输、交付、加工、分析、运维、运营及安全合规等数据任务的智能系统。其核心是围绕数据全生命周期构建闭环能力，通过对多源异构数据的感知、理解、处理与推理，将原始数据转化为可落地的洞察、决策与行动，是激活数据要素价值的核心技术载体。



图 2-1 数据智能体交互流程

（四）数据智能体核心能力特征

1. 数据环境感知能力

这是数据智能体的专属基础能力，可自主扫描企业全域多源异构的数据资产，动态感知表结构变更、数据血缘、质量状态与访问权限，构建实时更新的数据环境数字孪生，解决传统工具只能访问预定义数据的局限。

2. 数据业务指令理解与用户意图遵循能力

数据智能体具备的语义转换能力，能将自然语言中的业务术语、指标口径精准映射为数据库表、字段与计算逻辑，主动澄清模糊需求，从根源上解决“业务语言与数据语言不统一”的行业痛点。

3. 数据任务拆解与规划编排能力

专为复杂数据场景设计，可将宏观业务目标拆解为数据采集、清洗、分析等可执行子任务，生成数据处理有向无环图（DAG），并根据执行反馈动态调整策略，实现端到端数据任务自主推进。

4. 数据工具调用与全链路处理能力

内置数据领域原子技能库，可自主生成 SQL、Spark 等数据处理代码，调用数据库、清洗引擎、可视化平台等专属工具，一站式完成数据从原始形态到可用资产的全链路自动化处理。

5. 数据分析推理能力

突破通用智能体仅做浅层语义推理的局限，可基于全域数据开展描述、诊断、预测、规范全维度分析；通过因果推断等技术精准定位问题本质，输出可量化、可验证、可追溯的数据洞察与趋势预判，为业务决策提供科学依据。

6. 数据任务经验驱动的自主进化能力

从历史数据处理任务中沉淀专属知识，持续优化数据代码生成、任务规划与质量校验逻辑，自动适配新的数据模式与业务规则，无需频繁人工微调即可实现能力迭代升级。



图 2-2 数据智能体核心能力特征

（五）数据智能体带来的产业核心变革

数据智能体并非对传统数据工具的增量式改进，而是对数据要素利用全链路的底层范式重构。其依托数据感知、自主处理、高效协同、可信管控等能力，深度嵌入数据供给、流通、应用、治理全流程，精准破解传统体系数据供给能力薄弱、流通协同滞后、应用价值受限、安全治理缺位的核心瓶颈，推动数据要素利用模式从人工驱动向智能驱动全面转型，为数据要素价值释放提供核心技术支撑。

1. 数据供给模式变革：从静态规则驱动的低效供给向智能自治供给转型

传统数据供给环节过度依赖人工介入开展采集、加工与标注作业，面对多源异构数据场景适配性不足，原始数据向可用资产的转化效率偏低，难以支撑业务端持续、稳定的数据资源需求。数据智能体的出现，通过将智能感知与自动化处理能力赋予数据供给全环节，解决数据资源转化低效、供给能力不稳、异构数据适配困难的问题，实现数据从原始资源到标准化资产的自动化、高效率转化，筑牢数据要素高质量供给根基。

2. 数据流通机制变革：从孤岛式割裂向全域智能协同转型

传统数据要素在跨部门、跨主体流通中存在显著的链路壁垒，数据资源分散割裂且匹配对接效率低下，人工主导的流通模式极易造成流程阻滞，数据要素难以实现高效互通与协同利用。数据智能体的出现，通过将全域认知与智能协同能力赋予数据流通全链路，解决数据资源价值难以评估、分布分散割裂、流通链路阻滞、供需对接低效的问题，实现数据要素跨域顺畅流通与精准触达，打通数据价值流通的

关键枢纽。

3. 数据应用效能变革：从浅层被动支撑向深度主动赋能转型

传统数据应用多局限于固定规则下的基础统计分析，业务需求响应迟缓且价值挖掘深度不足，数据洞察速度慢，难以快速转化为实际决策支撑，无法适配实时化、精细化的业务运营需求。数据智能体的出现，通过自主分析与决策执行能力赋能数据应用全场景，解决数据应用过程中挖掘深度浅、响应效率低、价值转化不畅的问题，实现数据从被动辅助支撑向主动决策赋能升级，全面释放数据要素深度应用价值。

4. 数据安全治理变革：从事后被动处置向全程主动可控转型

传统数据治理与安全防护多采用事后核查的被动模式，缺乏对数据全流程的动态管控能力，数据质量隐患、合规风险与安全威胁难以提前预判，无法为数据要素利用提供可靠保障。从底层安全范式来看，传统体系普遍采用内嵌式、事后补救的设计思路，安全规则内嵌于业务系统内部，风险发现滞后、处置被动响应，且存在“决策与安全同体、既是选手又是裁判”的内生缺陷，已完全无法适配数据智能体自主决策带来的动态化、未知化新型风险。

二、数据智能体的形态及演化

（一）数据智能体与其他数据工具的差异

数据智能体与商业智能（BI）、机器人流程自动化（RPA）、自然语言转 SQL（NL2SQL）、通用大语言模型（LLM）的区别如表 1 所示。

表 1 数据工具区别

评估维度	商业智能 (BI)	机器人流程自动化 (RPA)	自然语言转 SQL (NL2SQL)	通用大语言模型 (LLM)	数据智能体 (Data Agent)
核心定位	数据展示与辅助分析工具	固定流程自动执行工具	自然语言查询转换工具	通用内容理解与生成模型	面向数据任务的自主分析与执行系统
驱动方式	指标体系、报表和预设模型驱动	规则、脚本和流程编排驱动	用户问题与数据库结构驱动	提示词和上下文驱动	业务目标、数据环境和反馈驱动
自主能力	较低，依赖人工选择分析路径	较低，按预设流程执行	较低，主要完成查询语句生成	中等，具备理解与推理能力	较高，可自主感知、规划、执行和验证
交互模式	仪表盘、报表和筛选操作	流程配置与任务触发	自然语言问数	多轮自然语言对话	多轮交互、任务执行与动态反馈
能力闭环	展示结果，通常不执行后续动作	完成固定流程，缺少自主调整	完成“提问—转换—查询—返回结果”	完成“输入—理解—生成”	形成“感知—规划—执行—验证—反馈”闭环
使用门槛	需要理解指标口径和报表结构	需要配置流程并持续维护	提问简单，但需明确数据口径	对话门槛低，依赖提示词设计	业务使用门槛较低，但建设需数据、工具和治理基础
容错自愈	依赖人工发现和修复问题	异常后按规则重试或转人工	可校验语法，复杂错误仍需人工处理	可通过反思修正，但结果稳定性有限	可依据执行反馈重试、切换工具、调整路径并请求人工接管
典型类比	标准化仪表盘和分析工作台	按流程执行的机器人	数据库查询翻译器	通用知识与内容助手	能使用工具、持续完成数据任务的数字员工

(二) 数据智能体主要形态

按照数据要素全生命周期的功能定位与价值分工，数据智能体可划分为供给类、流通类、应用类、安全合规类等四大核心形态，四类智能体沿数据价值链路各司其职、协同联动，共同构建覆盖数据生产、流通、应用、保障的全闭环能力体系。



图 2-3 数据智能体主要形态

1. 供给类数据智能体

供给类数据智能体聚焦数据生产端，核心目标是产出高质量数据资产，通过智能化手段完成数据采集、标注、加工与治理，打通原始数据向标准化资产转化的通道。

(1) 数据采集智能体

企业内部异构系统林立、外部数据源繁杂形成的“数据孤岛”长期难以破解，传统 API 对接成本高、ETL 管道维护难度大且易因源端变更断链，同时数据清洗依赖人工规则效率低下，跨系统数据语义对齐难、集成流程割裂，脏数据还会导致后续分析失真。数据采集智能体基于自然语言指令或目标库自动感知，智能匹配并生成数据抽取逻辑，一站式完成多源异构数据的自动化采集、智能清洗、标准化转换与跨源融合集成，可自动识别缺失值、异常值与重复数据，自主生成适配业务场景的清洗规则，实现数据语义对齐与去重合并，大幅降低了数据集成的工程门槛与运维成本，能够输出高质量、可直接用于分析的

标准化数据集。

(2) 数据标注智能体

大模型与行业算法训练过程中，人工标注成本高昂、效率低下，图文、音视频等多模态数据对齐困难，医疗、自动驾驶等垂直领域还存在专业知识门槛高导致的标注瓶颈。数据标注智能体依托大模型强大的学习能力，自主完成海量数据的预标注与交叉验证，结合主动学习机制仅将高价值边界数据交由人工审核，实现了训练数据产出效能的指数级跃升。

(3) 数据治理智能体

企业长期受“脏数据”困扰，传统数据治理存在元数据混乱、数据血缘断裂、质量规则依赖人工配置、问题发现滞后等管理困境。数据治理智能体专为数据资产全生命周期治理打造，驱动治理模式从“事后救火”转向“事前预防、事中自治”，可自动盘点全域黑盒数据资产、逆向解析完整数据血缘、基于历史规律自主生成并下发质量校验规则，自动识别并修复缺失值、异常值、逻辑冲突，实现了数据资产的自愈合。

2. 流通类数据智能体

流通类数据智能体聚焦数据流转环节，核心作用是破解供需错配与流通壁垒，实现数据资产的智能发现、高效传输与敏捷交付，畅通数据价值流转链路。

(1) 数据发现智能体

企业内部数据资产的分散导致“找数难、认数难”，跨部门、跨主体数据交易中存在严重的信息不对称，数据提供方“不知数据卖给谁、如何定价”，数据消费方“找不到所需数据、无法验证数据质量”，形成

双向供需错配。数据发现智能体作为数据资产的智能导航与撮合者，通过语义挖掘自动盘点全域数据资产、生成标准化资产画像与数据血缘图谱，支持自然语言精准检索与智能推荐，基于业务场景动态刻画供需双方画像，实现数据资产的精准匹配、智能定价辅助与自动化交易流通。

(2) 数据传输智能体

跨地域、跨云甚至跨国的海量数据流动，常遭遇网络拥塞、高延迟、带宽分配不均以及弱网环境下的数据丢包问题，影响数据流通的效率与可靠性。数据传输智能体深度融合网络拓扑与 AI 预测模型，能够感知全局网络状态，自主进行动态路由规划、智能带宽分配与协议调优，保障了高并发数据流通的低延迟、高吞吐与高可靠性。

(3) 数据交付智能体

传统模式下将底层数据打包为 API 接口或推送至下游业务系统，需要繁琐排期与定制化开发，成为数据工程的“最后一公里”障碍。数据交付智能体根据下游业务的调用诉求，自动组装数据逻辑，一键生成标准化数据 API 或数据包，并自主完成接口的限流、鉴权与发布配置，实现了数据资产向业务能力的敏捷转化。

(4) 数据加工智能体

传统多方数据融合加工高度依赖专业数据工程师跨源拉取数据、对齐口径规则并编写复杂加工逻辑，多源异构数据语义对齐难度大、标准不统一，定制化数据产品开发周期冗长，难以快速响应业务侧对衍生数据、特征数据集的需求，严重制约数据资产的价值转化效率。数据加工智能体能够理解业务人员的自然语言诉求，自主规划多源数据的关联融合与衍生计算逻辑，生成衍生指标、主题数据集、特征资

产等新型数据产品，显著提升数据产品的生产效率与数据资产的复用价值。

3. 应用类数据智能体

应用类数据智能体聚焦价值释放环节，核心作用是将数据资产转化为业务洞察与执行动作，支撑分析、科研、运营等场景的深度价值挖掘。

(1) 数据分析智能体

传统 BI 只能“预设维度、被动看数”，业务人员普遍缺乏数据分析技能，而专业数据分析师常年陷入临时性取数“泥潭”，导致数据价值难以快速转化为业务洞察。数据分析智能体化身“虚拟数字分析师”，通过思维链（CoT）将模糊的商业指令自主拆解为查询、统计算法与可视化图表生成，提供秒级的深度数据洞察与归因分析，真正实现了“人人都是数据分析师”。

(2) 数据运维智能体

以大语言模型或行业垂直模型为认知与决策核心，连接监控指标、运行日志、数据库、配置关系、业务拓扑、资产台账和历史案例等运维数据，自主完成环境感知、异常识别、根因分析、任务规划、工具调用、处置执行和结果验证的智能化系统。其典型场景包括自动巡检、告警降噪、故障诊断、容量评估、趋势预测、系统变更、故障自愈和运维报告生成，并通过知识库、长期记忆和反馈学习持续沉淀专家经验，形成“感知—分析—决策—执行—反馈”的运维闭环。

(3) 数据运营智能体

业务运营过程中，人工盯盘反应迟缓，海量异动指标无法实时归因，业务策略迭代往往滞后于市场变化（如广告投放、库存管理等场

景)，容易造成不必要的损失。数据运营智能体具备 7×24 小时全天候指标监控与感知能力，能够自主发现数据异常、深度挖掘根因，并直接调用底层业务系统工具，完成“洞察-决策-执行”的无缝闭环。

4. 安全合规类数据智能体

安全合规类数据智能体聚焦全链路保障，贯穿数据流转全流程提供合规管控与安全防护，确保数据活动全程合法合规、风险可控。

(1) 数据合规智能体

全球数据隐私法规日益严苛且动态更新，传统人工合规模式无法覆盖海量数据流转全链路，存在合规审核周期长、监管盲区多、成本高的核心痛点。数据合规智能体专为数据全生命周期合规管控打造，可自动解析最新法律法规并转化为可执行的机器规则，在数据采集、传输、存储、使用、销毁全环节实时扫描敏感信息，自动执行分级脱敏与权限管控，一键生成标准化合规审计报告与数据流转溯源记录，确保所有数据活动在法律框架内合规运行。

(2) 数据安全智能体

数据流通边界的扩大带来了内部越权、外部 APT 攻击、API 恶意滥用、数据批量泄露等各类安全威胁，传统被动防护模式难以有效应对这些新型风险。数据安全智能体专为数据流通全链路安全防护打造，基于用户实体行为分析动态构建个性化安全基线，可实时识别异常数据拉取、批量导出等高危行为并自动阻断，自主编排加密、脱敏、隐私计算等安全工具，在联邦学习、跨域数据共享场景中自动调度安全算力，构筑了数据流通的主动防御堡垒。

(三) 数据智能体能力演化

1. 形态演化的核心特征

从形态演化看，数据智能体并非传统数据工具的简单智能化升级，而是逐步实现由被动执行向主动感知、由单点操作向闭环协同、由静态功能向持续演进的转变。其核心特征主要体现在以下三个方面。

能力可成长：数据智能体突破传统工具功能随产品交付而基本固化的局限，依托长期记忆、知识沉淀和反馈学习机制，将历史任务中的业务规则、处理经验、异常案例和优化路径转化为可复用能力，并持续优化数据处理流程、分析方法、工具选择和风险控制策略。随着任务和反馈不断积累，智能体能够逐步适应数据结构、业务规则和使用需求的变化，减少对人工频繁配置规则和迭代功能的依赖。

行为可预期：数据智能体的任务规划、工具调用和执行行为受到企业业务规则、数据口径、权限边界和安全策略的共同约束。系统通过任务轨迹记录、数据来源标注、执行日志、结果校验和证据链留痕，使分析依据、决策过程和执行结果可解释、可审计、可追溯；对于超出权限、风险阈值或可信范围的任务，则触发人工确认或接管。清晰可控的能力边界，有助于降低人机协作中的信任成本，为智能体进入核心业务流程提供基础。

价值可共创：数据智能体逐步突破传统工具被动响应指令的定位，能够主动感知业务状态、识别数据异常、发现潜在问题并提出分析或处置建议。人类负责设定业务目标、提供专业经验、处理复杂判断并承担最终责任，智能体则发挥多源数据整合、持续分析、快速执行和经验复用等优势。双方通过任务反馈和结果修正不断优化协作方式，推动数据价值创造由“人使用工具”向“人机协同分析、决策与创新”演

进。

上述特征并非在数据智能体形成之初便完全具备，而是随着其自主感知、任务规划、闭环执行和自适应进化能力不断增强而逐级形成。基于这一演进逻辑，本文进一步将数据智能体划分为 L0 至 L5 六个能力成熟度等级，以反映其从数据工具、数据助手到首席数据官的能力跃迁。

2. 能力成熟度模型

依据自主感知、任务规划、执行闭环、自适应进化能力，数据智能体划分为 L0-L5 六个成熟度等级，每个等级对应一个直观角色定位，清晰体现能力层级跃迁。



图 2-4 能力成熟度模型

L0（固定指令执行级）：数据工具

仅能按照人类预先设定的公式、规则或操作步骤，完成单一的数据存储、计算、查询与展示任务，不具备自然语言理解、数据环境感

知、任务规划、工具选择和自主决策能力。数据源、处理方法、计算逻辑及结果判断均由人类负责，系统只承担被动执行功能，典型形态包括静态 ETL 脚本及固定格式 BI 报表等。

L1（辅助级）：数据助手

能够理解人类下达的明确任务指令，并根据任务需求自主选择和调用相应的数据工具，完成数据读取、查询、清洗、转换、汇总、分析及可视化等操作。人类负责提出具体目标并检查任务结果，智能体负责将自然语言指令转化为可执行步骤并调用工具完成任务，但不具备主动提出目标和独立进行业务决策的能力。

L2（部分自动化级）：数据工程师

具备特定业务场景下的数据判断、局部决策与自动执行能力，能够依据数据状态、预设目标、业务规则和权限边界，自主选择处置策略并调用相关工具或业务系统完成操作。例如，可根据数据质量检测结果选择修复方法，根据指标异常触发预警或工单，根据库存变化形成补货方案并执行标准化操作。该阶段实现了“判断—决策—执行”的局部闭环，但任务目标、决策范围和异常处置边界仍由人类预先设定。

L3（高度自动化级）：数据专家

具备端到端数据任务规划、跨工具协同与闭环执行能力，可在明确的数据、业务及授权边界内，实现“数据—洞察—判断—决策—执行”全链路自动化，并根据中间结果动态调整任务路径、处理异常和验证结果；超出权限、风险或可信边界时，主动请求人工接管。

L4（协作级）：数据总监

具备全域数据环境主动感知和复杂任务自主分解能力，可根据任务目标自主调度数据采集、加工、治理、分析、安全、合规与交付等

专项智能体协同工作，实现多智能体协同决策、闭环执行、异常处置与全过程质量保障。

L5（演进级）：首席数据官

具备持续自主学习与能力演进能力，可突破人类预先设定的具体任务流程和固定分析路径，自主从历史数据与多源数据中发现未知关系、潜在规律和风险信号，并在任务执行与反馈中持续探索更优决策、处理流程和资源配置方式，实现效率提升与成本优化。

三、数据智能体的技术体系

（一）数据智能体的参考架构



图 3-1 数据智能体参考架构

1. 数据业务语义对齐与意图解析模块

数据业务语义对齐与意图解析模块是数据智能体理解业务需求的入口，输入包括用户提出的自然语言问题，以及企业已有的业务术语、指标口径、数据地图和数据库信息。针对业务简称多、同名指标

口径不一致、业务语言与表字段难以对应等问题，采用业务本体、统一语义层、知识图谱、意图识别、实体识别、实体链接、槽位填充、语义消歧等技术，识别用户关注的业务对象、指标、维度、时间范围和筛选条件，并匹配对应的数据表、字段、关联关系与计算规则，最终输出结构化、口径明确的数据任务，为后续环境感知、任务分解和工具执行提供统一输入。

2. 动态全域数据环境感知模块

动态全域数据环境感知模块是数据智能体认识数据环境的“感知入口”，输入包括数据库、数据仓库、文件系统、业务接口及物联网设备中的数据结构、元数据、数据血缘、质量指标、权限信息和运行状态。针对企业数据源分散、结构变化频繁、数据关系复杂及实时状态难以掌握等问题，采用元数据采集、Schema 解析、数据目录、血缘解析、质量检测、变更监测、权限识别、状态监控、语义索引等技术，持续识别可用数据、关联关系及异常变化，最终输出动态更新的数据环境地图、数据资源清单和状态信息，为后续任务分解、技能选择、执行控制和结果验证提供依据。

3. 数据任务分解与编排模块

数据任务分解与编排模块是数据智能体的“任务规划中枢”，输入包括前序模块输出的结构化数据任务、数据环境地图、资源状态及权限边界。针对数据任务链路长、步骤依赖复杂、跨工具协同困难等问题，采用任务分解、有向无环图（DAG）、依赖分析、工作流编排、动态规划、并行调度、状态管理、异常重试、断点恢复、路径重规划等技术，将业务目标拆解为采集、清洗、转换、关联、建模、分析和交付等任务节点。最终输出包含节点关系、执行顺序、输入输出、调

用工具及异常处理规则的可执行任务计划，为数据技能驱动模块提供明确的执行路径。

4. 数据技能驱动模块

该模块是数据智能体将任务计划转化为实际操作的“执行引擎”，输入包括前序模块生成的可执行任务计划、数据对象、处理参数、资源状态及权限边界。针对数据工具分散、接口标准不一、能力难以复用以及模型无法直接操作底层系统等问题，采用原子技能（Skill）、专业插件（Plugin）、模型上下文协议（MCP）、工具注册、技能路由、参数映射、Text-to-SQL、Python、应用程序接口、数据库连接器、计算引擎适配器等技术，按任务节点调用数据库、数据仓库、计算集群和业务系统，完成查询、清洗、转换、计算、建模、分析与交付。最终输出数据结果、中间产物、执行状态和运行日志，并传递至结果验证与自主反思自愈模块。

5. 执行结果验证与自主反思自愈模块

该模块是数据智能体的“质量校验与纠错中枢”，输入包括数据技能驱动模块输出的数据结果、中间产物、执行日志、代码状态，以及预设的数据规则和性能指标。针对代码执行失败、数据口径偏差、结果不一致、计算异常和“数据幻觉”等问题，采用错误捕获、规则校验、Schema 校验、一致性检查、交叉验证、异常检测、单元测试、结果比对、置信度评估、反思推理、自动重试、回滚恢复、路径重规划、告警管理等技术，对任务过程与结果进行验证。最终输出通过校验的数据结果、问题诊断报告和修正后的执行方案；无法自主解决的问题触发告警或人工接管，并将验证结果和修正经验传递至数据知识沉淀与记忆进化模块。

6. 数据知识沉淀与长时记忆进化模块

该模块是数据智能体的“经验沉淀与能力进化中枢”，输入包括已验证的任务结果、执行轨迹、处理规则、异常案例、修正记录及用户反馈。针对业务经验分散、成功路径难以复用、同类问题重复发生以及能力依赖人工配置等问题，采用长时记忆、分层记忆、任务轨迹抽取、经验归纳、规则提炼、知识图谱、向量检索、检索增强生成、案例库、反馈学习、版本管理、记忆更新、遗忘机制等技术，将任务过程转化为结构化的业务规则、处理流程、典型案例和工具使用经验。最终输出可复用的知识、规则、标准作业流程和任务模板，并反馈至语义解析、任务规划、技能选择和结果验证等模块，使数据智能体能够在后续任务中调用历史经验并持续优化自身能力。

7. 安全控制模块

该模块是独立于主决策逻辑运行、贯穿数据智能体全流程的“安全与合规中枢”，输入包括用户指令、身份权限、任务计划、数据对象、工具调用请求、执行结果及运行日志。针对提示注入、越权访问、敏感数据泄露、高危操作失控和过程难以追溯等问题，采用指令校验、身份认证、角色权限控制、属性权限控制、最小权限、策略引擎、模型上下文协议权限管控、沙箱隔离、动态脱敏、敏感信息识别、密钥管理、内容过滤、风险分级、人工审批、全链路审计等技术，对数据访问、工具调用和结果交付进行实时控制。最终输出通过安全校验的执行指令和脱敏结果，并对高风险任务进行阻断、告警或转交人工处理，同时生成完整审计记录，保障数据智能体全过程安全、合规、可控、可追溯。

（二）数据智能体的关键技术体系

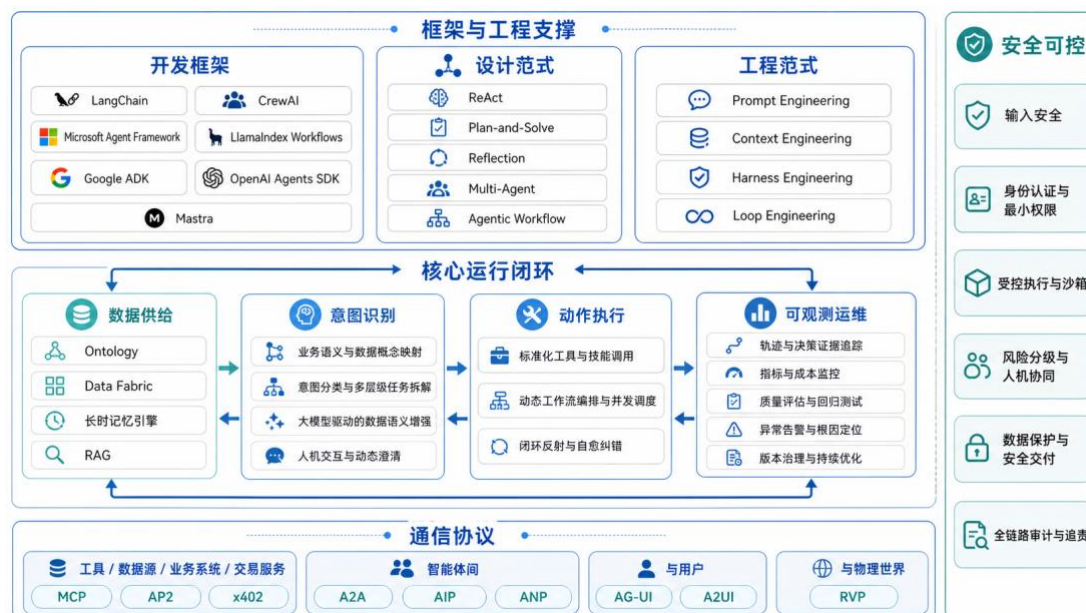


图 3-2 数据智能体关键技术

1. 开发框架：分层解耦、状态驱动与可观测的智能体工程框架

数据智能体开发框架是连接大语言模型、企业数据、业务知识、记忆系统和外部工具的工程底座，用于组织智能体完成需求理解、任务规划、工具调用、结果验证和持续优化。其核心作用是解决长链路任务中状态易丢失、工具调用不稳定、异常难恢复和运行过程难追踪等问题，并提供模型与数据接入、上下文与记忆管理、工作流编排、任务状态持久化、循环推理、人工介入、失败重试、断点恢复、链路追踪、成本监控和效果评估等能力。通过模块化、低耦合设计，模型、数据源、编排器和工具组件可独立替换与扩展，为数据处理、分析决策和业务执行提供稳定、可观测、可持续演进的运行环境。

典型的智能体开发框架包括 LangChain、CrewAI、Microsoft Agent Framework、LlamaIndex Workflows、Google Agent Development Kit

(ADK)、OpenAI Agents SDK 和 Mastra 等。LangChain 强调组件化集成, 并以 LangGraph 提供有状态编排、持久化执行和人工介入等能力; CrewAI 以 Crews 和 Flows 为核心, 支持角色化多智能体协作、任务分工及流程状态管理; Microsoft Agent Framework 融合 AutoGen 与 Semantic Kernel 的能力, 提供会话状态、类型安全、遥测和多智能体工作流编排; LlamaIndex Workflows 采用事件驱动和类型化步骤机制, 支持流式事件、状态持久化、人工介入及分步重试。Google Agent Development Kit (ADK) 提供多语言、代码优先的智能体构建、调试、评估和部署能力; OpenAI Agents SDK 采用少量核心抽象, 提供工具调用、任务转交、护栏、会话管理和执行追踪; Mastra 以 TypeScript 为核心, 集成智能体、工作流、持久化记忆、运行存储和可观测能力。企业应结合开发语言、云环境、任务复杂度和长期运维需求选择或组合使用, 不宜将整体架构绑定于单一框架。

2. 设计范式：用于决定智能体执行任务的底层逻辑路径

智能体设计范式是指构建智能体时, 为应对不同复杂度的任务而总结的通用架构模式与工作流。根据业界实践, 智能体设计范式主要分为以下五大类, 它们既可以独立使用, 也可以组合形成更复杂的智能体系统。

(1) 基础交互范式：ReAct（推理与行动结合）

ReAct 通过“推理—行动—观察”循环, 将模型判断与工具调用交替进行。智能体分析当前任务后, 选择搜索、数据库或应用程序接口等工具, 并依据返回结果继续推理和修正, 直至完成任务。该范式适用于智能问数、日志分析、异常排查和动态检索等信息不完整、路径

难预设的场景，具有灵活、过程易追踪等优势，但长任务中容易出现重复调用、路径偏移和无效循环，并增加响应时间与运行成本。

(2) 规划执行范式：Plan-and-Solve（先规划后执行）

Plan-and-Solve 将任务规划与具体执行相分离，先从全局角度拆解目标，再按照计划逐步完成。智能体将复杂任务划分为具有顺序和依赖关系的子任务，明确所需数据、工具及预期结果，并在执行偏差时进行调整或重新规划。该范式适用于数据加工、复杂分析、报告生成和代码开发等结构较明确的任务，具有路径清晰、执行稳定和便于检查等优势，但初始规划中的遗漏或错误可能向后传递，因此对规划质量和过程校验要求较高。

(3) 反思优化范式：Reflection（反思与迭代优化）

Reflection 通过“执行—反思—优化”循环，使智能体能够检查并修正初始结果中的逻辑漏洞、事实错误、格式问题和内容遗漏。智能体生成初步结果后，由自身或独立评审模块依据预设标准进行检查，形成修改意见并重新生成或局部修正，直至达到质量要求。该范式适用于代码调试、分析校验、报告撰写和方案评审等高质量场景，能够提升结果准确性与完整性，但会增加计算成本和响应时间，并需设置明确的评审标准与迭代上限。

(4) 多智能体协作范式：Multi-Agent（多智能体协作）

Multi-Agent 通过角色分工和协同机制，将复杂任务分配给多个具备不同职责和专业能力的智能体。系统可设置规划者、执行者、分析者、评审者和汇总者，由协调智能体完成任务分配、状态同步和结果整合。该范式适用于复杂数据分析、跨部门业务处理、软件研发和大型调研等跨领域任务，可发挥专业分工、并行处理和交叉验证优势；

但系统结构较复杂，需要解决通信成本、任务冲突、信息重复、状态一致性和结果整合等问题。

(5) 综合工作流范式：Agentic Workflow（智能体工作流）

Agentic Workflow 将任务规划、工具执行、结果反思、多智能体协作和人工介入整合为统一工作流。系统围绕业务目标自主推进任务，根据执行结果和环境变化动态调整路径，并对异常步骤进行重试、回滚或重新规划；高风险、低置信度及越权操作则转交人工审核。该范式适用于数据流水线、经营分析、客户服务和软件开发等长周期、多系统的企业场景，能够实现端到端自动化，但对状态管理、权限控制、异常恢复、成本管控和质量评估要求较高。

上述范式并非相互替代，而应根据任务复杂度、流程确定性、质量要求、风险等级和成本预算灵活组合。信息不完整、需要动态检索与结果校验的任务，可采用“ReAct + Reflection”；步骤较多且分工明确的复杂任务，可采用“Plan-and-Solve + Multi-Agent”。企业级数据智能体还应叠加状态持久化、人工介入、安全审计、失败恢复和成本控制机制，使任务在保持自主性的同时具备稳定、可控和可追溯的工程运行能力。

3. 工程范式：用于保障智能体稳定执行持续优化的工程方法体系

智能体工程范式关注如何将大语言模型、企业数据、业务知识、工具和系统组织为稳定运行的工程体系，重点解决模型按照什么规则工作、依据哪些信息判断、如何安全调用工具以及怎样根据反馈持续改进等问题。其主要包括提示词工程、上下文工程、驾驭工程和循环工程。四者分别负责行为约束、信息供给、执行管控和反馈收敛，共

同支撑数据智能体由模型原型走向可持续运行的生产级应用。

(1) 提示词工程 (Prompt Engineering)

提示词工程用于明确智能体的角色、目标、规则和输出要求，通过系统提示词、任务说明、示例、格式约束和工具调用说明，引导模型按照预期方式处理任务。在数据场景中，应明确业务口径、分析步骤、数据边界、异常处理方式和结果格式，使模型稳定完成意图识别、任务分解、代码生成和结果解释。提示词工程是控制模型行为的基础，但主要适用于单次或相对固定的任务，难以独立解决动态环境感知、长任务状态管理和复杂工具协同问题。

(2) 上下文工程 (Context Engineering)

上下文工程用于为智能体动态组织完成当前任务所需的信息，通过检索、筛选、排序、压缩和组装，将业务本体、指标口径、数据库结构、数据血缘、历史任务、知识文档、工具说明、权限边界和执行状态提供给模型。其目的是在有限窗口中选择最相关、最新且可信的内容，减少信息缺失、冲突、过载和状态丢失。该工程连接企业数据底座与模型推理能力，使智能体能够依据真实业务规则和数据环境作出判断。

(3) 驾驭工程 (Harness Engineering)

驾驭工程围绕大语言模型构建执行与管控环境，将模型生成的计划转化为安全、稳定、可观测的实际操作。其通过上下文组织、工具与技能注册、能力路由、工作流编排、状态持久化、权限控制、沙箱隔离、失败恢复、人工审批和运行追踪等机制，连接数据库、计算集群、知识库与业务系统，并可接入 MCP 工具协议和效果评估体系。该工程明确模型能够访问哪些资源、如何调用工具、执行过程如何受

控以及失败后如何处理,使大语言模型具备执行真实任务和交付业务结果的工程能力。

(4) 循环工程 (Loop Engineering)

循环工程用于设计和管理模型、工具与运行环境之间的持续交互,使智能体在长链路任务中逐步完成目标。其基本过程可概括为“目标理解与状态感知—推理与规划—工具执行—环境反馈—结果判断—调整或终止”,并可根据任务需要加入结果验证、反思修正和路径重规划。系统依据工具返回结果、数据质量、执行状态和用户反馈,采取自动重试、参数调整、工具切换、路径重规划、结果回滚或人工接管等措施。同时设置迭代次数、时间预算、成本上限、置信度阈值和终止条件,防止无效循环。在接入长期记忆、轨迹分析和反馈学习机制后,执行经验还可用于优化后续任务路径。

4. 数据供给：用于保障底层数据与业务知识的精准供给

(1) 本体 (Ontology)

保障智能体输出确定性的核心工程化底座,专门解决大模型概率性输出导致的“幻觉”与业务逻辑偏差问题。它通过将企业统一的业务规则、指标口径、实体关系与行业知识映射为结构化的数字本体网络,为大模型的推理过程施加刚性约束,确保所有分析结果都符合企业业务常识,大幅降低人工校验成本,是数据智能体落地企业的必备基础。

(2) 数据编织 (Data Fabric)

构建全域统一的逻辑数据底座,破解企业内部异构系统林立、物理数据分散导致的“数据孤岛”顽疾。它利用主动元数据管理技术,自动发现、映射与关联分散在数据库、API、文档中的各类数据,将物理上分散的数据编织成逻辑上统一的透明网络,让智能体能够像访问

本地数据一样无缝访问全域数据，无需关心数据的物理存储位置与格式。

(3) 长时记忆引擎与 RAG

赋予智能体专业业务背景的核心组件，系统性解决大模型知识过时、领域知识不足与上下文窗口有限三大痛点。长时记忆引擎负责沉淀智能体的历史交互经验与任务执行记录，实现持续学习；RAG 则通过向量检索技术，实时挂载内部知识库、业务文档与最新数据，为大模型提供精准的上下文增强，显著提升输出的准确性与时效性。

5. 意图识别：用于精准解析意图并约束大模型的推理逻辑

数据智能体中的意图识别并非简单判断用户“想查什么”，而是要同时理解业务含义、数据口径和任务目标。面对“最近业务怎么样”“为什么利润下降”等模糊、隐喻或多步骤问题，系统需要将自然语言与企业数据资产准确对应，并明确查询范围、分析方法和执行边界，最终把业务问题转化为结构化、可执行、可确认的数据任务。

(1) 业务语义与数据概念映射

该环节从用户问题中识别指标、维度、时间范围、筛选条件和业务对象，并通过数据字典、业务本体、知识图谱和 Schema 检索，将“销售额”“大客户”等业务表达映射至具体表、字段、指标口径和计算规则。同时结合多轮对话、当前页面及用户权限补全省略信息、消除口径歧义，输出统一的业务语义和数据对象，为后续任务拆解提供准确基础。

(2) 意图分类与多层级任务拆解

系统根据任务复杂度，将用户意图区分为指标查询、归因诊断、趋势预测和动作执行等类型。对于复杂需求，则通过任务规划和分步

推理，将抽象目标拆解为数据提取、异常定位、维度下钻、因素分析和结果交付等子任务，明确各步骤的依赖关系、所需工具及预期结果，形成结构化执行计划，避免复杂分析停留在单次查询或表面回答层面。

(3) 大模型驱动的数据语义增强

为避免自然语言直接生成结构化查询语言造成口径偏差和数据幻觉，系统先将用户意图转换为 JSON、DSL 或语义树等结构化中间表示，明确指标、维度、筛选条件和计算逻辑。随后通过 Schema-RAG 动态注入相关表结构、枚举值、数据血缘和近义词，并结合行业示例进行上下文学习，使业务语义与具体数据库语法解耦，提高意图解析的准确性、稳定性和可复核性。

(4) 人机交互与动态澄清机制

当系统识别到多个指标口径、数据来源或执行路径，或意图置信度低于设定阈值时，应主动向用户提出澄清选项，而不是直接猜测。对于高耗时、高成本或高风险任务，还需在执行前回传意图理解、数据范围和分析口径，由用户确认后再触发执行。用户的纠正结果和常用口径可进一步写入长期记忆，持续提升个性化意图识别能力。

6. 动作执行：用于驱动外部工具完成数据任务的实际落地

动作执行是将意图理解和任务规划转化为真实操作的关键环节。执行引擎根据任务计划选择相应技能和工具，连接数据库、计算平台、算法服务及业务系统，完成数据查询、加工、分析、结果交付和业务操作，并通过任务调度、状态管理和异常恢复保障执行过程稳定、可控。

(1) 标准化工具与技能调用

标准化工具调用用于将智能体生成的动作计划转换为外部系统

能够识别和执行的指令。系统通过应用程序接口、模型上下文协议、数据库连接器和专业插件接入数据平台及业务系统，并以 Skill 形式封装查询、清洗、分析、绘图、消息发送和状态更新等可复用能力。执行时，智能体根据任务节点选择相应 Skill，完成参数映射、身份鉴权和调用控制，精准触发结构化查询语言、Python 程序、算法服务或业务接口，并返回统一格式的执行结果和状态信息。

(2) 动态 workflow 编排与并发调度

针对步骤多、依赖关系复杂的长链路任务，执行引擎通过动态 workflow 实现任务拆解、资源调度和过程控制。系统将任务组织为有向无环图，自动识别节点之间的先后依赖与并行关系：相互独立的数据查询和接口调用可异步并发执行，存在依赖的任务则在前置结果完成后继续推进。同时，执行引擎持续记录节点状态、中间结果和资源占用情况，并根据执行进度动态调整任务顺序和并发规模，缩短整体处理时间，避免重复计算和资源冲突。

(3) 闭环反射与自愈纠错

执行引擎通过结果验证和错误反馈形成“执行—校验—修正—重试”的自愈闭环。在任务完成前，系统利用规则约束、合理性检查和结果比对，识别空数据、指标越界、口径冲突等异常；当结构化查询语言或 Python 程序执行失败时，自动捕获错误信息和异常堆栈，并将原始代码、数据结构及报错信息重新提交模型进行修正。系统同时保存变量、DataFrame 和阶段结果，支持有状态分步调试、断点恢复和局部重试；无法自动解决的问题则触发告警或人工接管。

7. 通信协议：用于统一通信标准以支撑多智能体与跨域协同

通信协议是数据智能体技术体系的核心支撑层，构成了智能体生

态的“通用语言”，系统性解决了工具调用、内部协同与跨域互联三大基础问题。

(1) 智能体与外部工具、数据源、业务系统及交易服务之间

该层解决智能体如何连接外部系统并完成实际操作的问题。MCP（模型上下文协议）统一智能体访问文件、数据库、搜索服务和企业系统的接口，将外部能力封装为工具、资源和提示模板，减少重复开发。涉及付费服务时，AP2 通过意图声明、授权凭证和交易记录，明确用户授权、交易金额和责任边界；x402 则面向应用程序接口、数字内容等场景，支持按次付费和自动结算。三者分别解决能力接入、交易授权和支付结算问题，使智能体从“调用工具”进一步走向商业服务闭环。

(2) 智能体与智能体之间

该层解决不同智能体如何发现彼此、分配任务、同步状态和交付结果的问题。A2A（智能体间协议）侧重任务协作，通过智能体名片说明能力和访问方式，并以任务、消息、状态和交付物为核心对象，实现跨平台智能体之间的委托、跟踪与结果返回。AIP（智能体互联协议）提供“身份—描述—发现—交互—工具调用”的统一规则，支持可信接入、能力匹配、同步或异步协作及行为审计。ANP（智能体网络协议）面向开放网络，通过去中心化身份、加密通信、能力发现和协议协商，使智能体能够跨组织、跨域名和跨平台安全互联。三者分别支撑任务协作、标准化互联和开放网络协同。

(3) 智能体与用户之间

该层解决智能体如何向用户呈现任务过程、接收用户操作并提供可交互界面的问题。AG-UI（智能体—用户交互协议）侧重智能体运

行时与前端应用之间的双向事件通信，将任务状态、消息内容、工具调用、中间结果和最终结果同步至用户界面，并支持用户确认、暂停、干预和继续执行，使智能体运行过程可见、可控，前后端状态保持一致。A2UI（Agent to UI，智能体到用户界面协议）侧重界面的声明式动态生成，智能体在客户端允许的组件目录范围内，输出按钮、表单、图表等结构化界面及数据描述，由客户端渲染为可操作的原生界面，并将用户选择、填写和提交等操作反馈给智能体。二者分别支撑“智能体与前端应用的运行通信”和“智能体驱动的动态界面生成”，可以组合使用，共同推动人机交互由纯文本问答向过程可视、状态可同步、操作可干预的应用形态演进。

(4) 智能体与物理世界之间

该层解决数字智能体如何与机器人、物联网设备、数字孪生等物理实体建立连接、交换状态并协同完成任务的问题。以 RVP（虚实融合智能体通信协议）为例，其侧重数字世界与物理世界之间的动态交互，通过统一的实体标识、任务消息、状态反馈和协同机制，支持智能体向物理设备下发任务、获取运行状态、接收执行结果并开展协同调度。虚实融合智能体通信协议可与 MCP 等工具接入机制以及 A2A、ANP 等智能体协作机制配合，分别支撑工具连接、智能体协同和物理执行，形成“数据与工具—数字智能体—物理实体”的交互链路，适用于智能制造、机器人协同、城市治理和物联网调度等虚实融合场景。

8. 安全可控：用于构筑风险防线并确保智能体行为绝对可控

(1) 指令与输入安全控制

对用户指令、外部文档、网页内容和工具返回信息进行安全检测，识别提示词注入、恶意指令、越权意图和敏感信息，防止智能体被非

可信内容诱导改变任务目标或泄露数据。主要技术包括指令校验、内容过滤、提示词注入检测、任务边界约束和风险意图识别。

(2) 身份认证与最小权限控制

根据用户身份、岗位职责、任务目标和数据敏感等级，动态限定智能体可以访问的数据、工具和操作范围。智能体仅获得完成当前任务所需的临时权限，任务结束后自动收回，避免长期授权、权限冗余和跨场景越权。主要技术包括身份认证、角色权限控制、属性权限控制、动态授权和凭证管理。

(3) 受控执行与沙箱隔离

对 Python、结构化查询语言、Shell 脚本及其他可执行操作建立隔离运行环境，限制网络连接、文件读写、系统调用、运行时间和资源消耗。执行前进行代码和参数校验，执行中持续监控异常行为，执行后保留日志和结果。沙箱能够限制错误或恶意代码的影响范围，但仍需与权限控制、数据脱敏和审计机制配合使用。

(4) 风险分级与人机协同控制

按照操作对象、数据敏感度、影响范围和结果可逆性，对任务进行风险分级。低风险查询可自动执行，中风险操作需用户二次确认，高风险操作如批量导出、永久删除、修改核心规则和生产系统变更，则必须经过人工审批。对于低置信度、超出权限或执行异常的任务，系统自动暂停并转交人工处理，确保关键决策始终保留人类最终控制权。

(5) 数据保护与安全交付

在数据读取、处理和输出过程中实施敏感信息识别、动态脱敏、结果水印、下载控制和到期销毁，防止智能体在回答、报告或文件交

付中泄露敏感信息。

(6) 全链路审计与责任追溯

完整记录用户指令、模型判断、数据访问、工具调用、人工审批、执行结果和异常处置过程，使每项操作都能够回答“谁发起、为何执行、访问了什么、产生了什么结果”，为安全复核、责任界定和合规审计提供依据。

9. 可观测运维：用于实现全链路追踪、质量评估与持续运营

数据智能体运行过程涉及模型推理、任务规划、工具调用、数据访问和多轮反馈，传统应用监控难以解释任务为何失败、结果是否可信以及成本消耗在哪个环节。因此，可观测运维需要围绕“执行过程、结果质量、运行状态和版本变化”建立统一监测体系，使问题可发现、原因可定位、效果可评估、变更可回退。

(1) 全链路执行轨迹与决策证据追踪

该能力用于记录智能体从接收任务到交付结果的完整执行过程，包括用户需求、意图解析结果、任务计划、节点状态、模型与提示词版本、工具调用参数、数据来源、中间结果、异常信息和人工介入记录。系统通过统一任务标识串联模型、工具和业务系统日志，形成可视化执行链路。出现结果错误、工具调用失败或任务循环时，运维人员可定位具体问题节点。追踪重点应是可验证的执行事实和决策依据，而非保存或展示模型未经整理的内部思维过程；同时还应对敏感字段进行脱敏，并设置日志保存期限和访问权限。

(2) 运行指标与资源成本监控

该能力用于持续监测智能体的运行效率和资源消耗，指标包括任务成功率、响应时间、工具调用成功率、重试次数、循环深度、人工

接管率、模型调用量、Token 消耗和单任务成本等。系统可按智能体、业务场景、模型、工具和用户进行分层统计，识别高延迟、高失败率和高成本环节，并通过模型路由、缓存复用、并发控制和调用限额优化运行效率。

(3) 结果质量评估与回归测试

仅监测系统是否正常运行，不能判断智能体输出是否正确。该能力通过规则校验、标准答案集、人工抽检、模型评审和业务指标反馈，对结果的准确性、完整性、数据一致性、依据充分性和格式合规性进行评估。对于提示词、模型、知识库、技能或工作流的每次调整，均应进行离线回归测试和版本对比，避免局部优化导致其他场景效果下降。

(4) 异常告警与根因定位

系统根据任务失败率、响应时间、成本变化和质量指标设置动态阈值，发现模型不可用、工具异常、数据源变更、任务循环和输出质量下降等问题时，及时触发告警。通过关联任务轨迹、模型调用、工具日志、数据血缘和系统资源，定位问题来自模型、提示词、数据、工具还是工作流，并采取降级模型、切换工具、限制调用、暂停任务或人工接管等措施。

(5) 版本治理与持续优化

对模型、提示词、知识库、技能、工具接口和工作流进行统一版本管理，完整记录配置变更、评测结果和发布范围。优化方案应经过离线评估、人工审核、灰度发布和效果验证后再逐步上线，出现异常时支持快速回滚。历史运行数据可用于发现高频失败模式和流程瓶颈，但不宜由智能体未经验证直接修改生产环境中的提示词或路由策略。

四、数据智能体的产业生态

（一）产业生态全景图谱



图 4-1 数据智能体产业生态

1. 基础底座

基础底座是数据智能体运行的资源与能力基础，主要包括算力基础设施、基础大模型和数据基础设施。算力基础设施提供模型训练、推理和任务并发所需的计算、存储与网络资源，并支持公有云、私有化和混合部署；基础大模型承担自然语言理解、任务规划、代码生成和分析推理等核心能力，也可结合行业数据形成专用模型；数据基础设施负责连接数据库、文件、知识库和业务系统，统一管理结构化与非结构化数据，为智能体提供准确、实时、可调用的数据来源。三者共同决定数据智能体的性能、成本、安全性和后续扩展能力。

2. 开发框架和工具

开发框架与工具是连接基础底座与上层应用的能力中枢，主要包括智能体编排与调度框架、工具调用协议与组件、记忆与检索增强组

件。编排与调度框架负责拆解任务、安排执行顺序并协调多个智能体协同；工具调用协议与组件用于统一接入数据库、业务系统和外部服务，使智能体能够真正执行查询、计算和操作；记忆与检索增强组件则负责调用企业知识和历史经验，提高回答准确性与任务连续性。三者共同决定数据智能体能否稳定执行复杂任务，并实现能力复用和规模化扩展。

3. 开发平台

开发平台是面向企业和开发者的一站式智能体建设与部署载体，整合模型接入、知识库、工具调用、流程编排、测试评估、发布运维等能力，加快数据智能体从方案设计走向业务应用。通用智能体开发平台提供较完整的开发工具链，适用于企业自主建设多类数据智能体，如阿里云相关平台；低代码 / 无代码搭建平台通过拖拽编排和模板配置降低开发门槛，如火山引擎、Dify；智能体能力开放平台则通过应用程序接口、软件开发工具包、Skill 和 MCP 市场开放模型、数据、工具及行业能力，支持快速接入现有业务系统。三类平台共同推动智能体建设由定制开发向标准化、组件化和规模化演进。

4. 场景解决方案

场景解决方案是数据智能体产业价值落地的核心环节，围绕数据从生产、流通到应用与保障的全生命周期形成多类专业智能体。数据采集、标注和治理智能体主要面向数据生产端，分别完成多源数据接入与清洗、多模态数据预标注与质检，以及坏数据识别、补全修复和质量提升；数据供需、传输、交付和加工智能体聚焦数据流通，通过资产发现与供需匹配、智能路由与带宽调度、标准接口与数据包生成，以及多源融合和衍生计算，推动数据安全高效流转并转化为可复用的

数据产品。数据分析、运维和运营智能体面向业务应用，承担自然语言问数、归因预测、系统巡检、故障诊断、指标监控和策略执行等任务，形成“数据—洞察—决策—行动”闭环。数据合规和安全智能体则贯穿全过程，负责法规识别、敏感数据管控、异常行为监测和风险处置，保障数据活动合法、合规、安全、可追溯。

5. 开源与社区

开源与社区是数据智能体能力复用和生态扩展的重要支撑，主要包括 Skill 商店、MCP 市场和智能体市场。Skill 商店将数据查询、清洗、分析、绘图和报告生成等专业能力封装为可复用技能，降低重复开发成本；MCP 市场集中提供数据库、文件系统、办公软件和业务系统等标准化连接组件，使智能体能够快速接入外部工具与数据源；智能体市场则面向具体岗位和业务场景，提供可直接使用或二次配置的智能体产品。三类市场共同形成“技能复用—工具连接—应用交付”的生态链路，推动数据智能体由单体开发走向开放协作和规模化应用。

6. 服务保障

服务保障覆盖数据智能体从规划建设到持续运营的全过程，主要包括咨询与实施、运维与运营、评测与标准三类服务。咨询与实施服务负责需求诊断、方案设计、系统集成和部署交付；运维与运营服务对模型、工具调用、工作流、成本及用户反馈进行持续监测和优化；评测与标准服务则围绕准确性、稳定性、安全性和业务效果建立评价与规范体系，共同保障数据智能体顺利落地、稳定运行和持续改进。

（二）商业模式探索

1. 按量计费

按量计费是数据智能体最灵活的付费模式，核心是“用多少付多少”，无前期固定投入，完全按实际资源消耗结算。计费维度通常涵盖智能体调用次数、Token 消耗量、数据处理量、计算资源占用时长等，支持按需弹性扩容。该模式特别适合中小企业、初创团队以及临时性、突发性的数据处理需求，能有效降低试错成本。

2. 订阅制 SaaS 模式

订阅制 SaaS 是当前数据智能体最主流的商业模式，用户按月度或年度支付固定费用，即可获得标准化的产品功能与持续的服务支持。服务内容通常包含基础数据智能体能力、定期版本更新、7×24 小时技术支持以及一定额度的资源配额。该模式成本可预测、无需本地部署与运维，适合成长型企业的常态化数据需求。

3. 独立部署

独立部署模式是将数据智能体系统完整部署在企业本地服务器或私有云环境中，所有数据与计算过程均在企业内部完成，数据不出域。该模式核心优势是极致的数据安全性与隐私保护，同时支持深度定制化开发与性能调优，能与企业现有 IT 基础设施无缝集成。主要面向金融、政务、能源等对数据安全有严格要求的行业客户。

4. 定制化解决方案

定制化解决方案是针对大型企业或特定行业的复杂业务需求，提供从需求调研、方案设计、系统开发到部署运维的全流程端到端服务。该模式深度融合企业的业务流程、数据体系与管理规范，打造专属的数据智能体系统，解决企业核心痛点。通常包含前期咨询、定制开发、上线培训与长期运维等环节，项目周期较长但适配性最强。

五、数据智能体的场景应用

（一）数据智能供给与治理

当前，数据加工与供给仍面临处理链路割裂、脚本维护成本高、标注智能化程度较低、专业人员不足、质量控制困难以及生产周期较长等问题，难以满足多模态数据处理和高质量数据集规模化供给的需求。针对上述问题，数据智能体通过融合大语言模型、视觉模型、知识图谱和专业数据工具，逐步承担数据解析、清洗转换、智能预标注、数据增强、异常检测和质量评估等重复性任务，同时保留人工在规则制定、疑难样本审核和最终质量验收等关键环节的参与，推动数据生产由人工密集型作业向“智能处理—人工复核—质量检验—反馈优化”的人机协同模式转变。

伴随智能体能力不断嵌入数据生产流程，数据智能加工与供给已由单一的数据清洗或辅助标注工具，逐步扩展为涵盖多源数据接入、元数据探查、格式识别、清洗转换、智能标注、语义补全、质量检查、数据集管理和反馈优化的综合能力体系。当前产业实践主要沿两条路径推进：一类面向通用结构化数据与多模态数据，通过自然语言驱动任务创建、自动生成数据结构、调用数据处理工具和组织多角色智能体协同，并结合本地化部署、过程监控和结果追溯，形成覆盖“采集—加工—标注—质检—交付”的一体化数据生产平台；另一类面向医学影像、自动驾驶、遥感、能源、工业、政务和小语种等专业场景，将行业模型、知识图谱、领域标准、标签体系和专家经验嵌入数据生产流程，通过智能预标注、数据增强、自动质检、人工精标和反馈学习相结合的方式，提高垂直领域数据加工的效率、一致性和专业性。

总体来看，现阶段数据智能体并非完全替代人工标注与质量审核，而是逐步承担任务拆解、数据探查、预处理、预标注、争议样本筛选、质量检查和经验沉淀等工作，使人工更加聚焦于专业判断和质量把关。

Google 团队研发的 Adjudicator 系统针对标签噪声难识别、单一大模型易产生幻觉和判断依据不足等问题，融合大语言模型、动态知识图谱与多智能体，通过证据检索、协同审议和加权投票完成标签纠错。在 1000 条平衡测试样本中，系统取得 0.99 的 F1 值，体现了多智能体与结构化证据在高精度数据验证中的潜力。

鹏城实验室将数据智能技术应用于 PCB 柔性制造，针对多品种、小批量生产过程中工艺数据分散、设备数据异构、工步关联关系复杂、工艺知识依赖人工经验等问题，系统以工业数据治理和国产大模型为核心，对设备、工艺、工步及排产等多源数据进行融合治理，构建工步级工业数据模型，通过工艺知识学习和智能调度策略生成，实现复杂工艺约束理解、动态排产优化和智能辅助决策。应用后，订单平均超时天数缩短 60% 以上，瓶颈工序平均滞留时间缩短 5% 以上，支撑特殊工艺产品生产周期缩短 5% 的目标，形成面向离散制造的数据治理与智能决策能力。

蚂蚁集团将相关能力应用于规模化数据生产，建设数据标注全链路智能体助手。针对标注任务规模大、专业性强、规则多变及人工标注、质检效率不足等问题，系统以大语言模型和智能体框架为核心，通过 skill 文档封装专家知识、scripts 固化流程，并由路由、标注和质检智能体协同完成规则校验、知识检索、自动标注/。应用后，人工标注工时平均节省近 30%，智能化覆盖率稳定在 30% 至 40%，同时提升了任务流转效率和运营一致性。

（二）数据智能加工与交付

长期以来,企业数据分散在不同业务系统、数据库和文件平台中,普遍存在数据目录不完整、字段含义不清、资产关系不明、数据标准不统一以及跨系统接口不一致等问题,导致“找数难、认数难、加工难、用数难”。尤其在多源数据融合加工过程中,往往需要人工完成数据抽取、口径对齐、清洗转换、关联匹配和衍生计算,再将结果封装为数据集、报表或接口,流程长、重复开发多且难以快速响应业务需求。针对上述痛点,数据智能体通过元数据探查、语义理解、自然语言检索、任务规划和工具调用,将数据发现、需求匹配、多源融合加工与结果交付衔接起来,推动数据服务由人工查找和定制开发,向“智能发现—融合加工—按需组织—安全交付—结果反馈”的协同模式演进。

从当前产业实践看,数据智能加工与供给已由单一的数据目录查询或接口调用,逐步扩展为涵盖多源数据接入、结构识别、元数据采集、语义补全、资产画像、自然语言找数、跨源关联、清洗转换、指标计算、数据产品生成和交付留痕的综合能力体系。部分产品能够自动识别数据库及表、字段、分区和血缘关系,建立持续更新的数据目录;部分产品则根据业务需求,自主规划不同数据源之间的关联、融合与衍生计算逻辑,生成主题数据集、衍生指标、特征资产、报表或标准接口。现阶段,数据智能体主要承担需求理解、数据定位、加工流程编排、工具调用、状态跟踪和结果推送等工作,而数据口径、访问权限及高风险操作仍需依托既有平台规则并保留人工审核,从而兼顾加工效率、交付质量、数据安全与过程可追溯。

南京市计量监督检测院建设的 AI 计量数据交互智能体，针对企业计量管理系统与计量服务平台数据链路割裂、重复录入和跨平台操作繁琐等问题，通过标准化接口、异构系统连接、数据双向同步和状态自动回传，贯通报检、预约、进度跟踪、证书获取及到期预警等环节。企业无需改造原有系统，即可在自有系统内完成业务办理并实时获取计量数据，有效降低人工维护成本，提升业务效率及数据交付的安全性及可追溯性。

（三）数据智能分析与运维

传统数据分析长期存在业务人员缺乏专业工具使用能力、数据分析师重复取数任务繁重、指标口径容易偏差以及复杂分析响应周期较长等问题；当数据结果出现异常时，又需要数据研发人员在跨集群、多租户和长血缘链路中反复核查数据、代码与加工逻辑，问题定位高度依赖个人经验。针对上述痛点，数据智能体融合自然语言理解、语义建模、知识图谱、多智能体协同、工具调用、机器学习和长期记忆等技术，将模糊业务问题转化为数据查询、分析计算、归因判断和报告生成任务，并进一步衔接异常发现、根因追溯、方案建议和经验沉淀，推动数据工作由人工取数和被动排查向“智能分析—结果校验—异常诊断—辅助决策”的协同模式转变。

当前，数据智能分析与运维已由简单的自然语言转结构化查询，逐步扩展为涵盖多源异构数据接入、业务语义理解、指标口径约束、分析计划生成、代码执行、异常检测、归因分析、趋势预测、可视化报告和运维诊断的综合能力体系。在分析环节，相关产品通过语义层、本体、知识图谱和结构化逻辑表示，将业务问题转化为可执行、可复

核的分析逻辑，并利用多智能体规划、反思和交叉验证提高结果可靠性；在运维环节，则融合监控指标、日志、配置关系、数据血缘、历史案例和专家知识，辅助完成风险预警、异常识别、根因定位和处置建议。同时，通过权限继承、私有化部署、代码沙箱、人工审批和全链路审计，保障分析与执行过程安全可控。现阶段，数据智能体主要承担重复取数、标准化分析和常规问题排查，复杂业务判断及高风险操作仍需专业人员参与。

在数据运维方面，蚂蚁集团建设的 Eva 数据问题排查助手，针对离线数仓规模庞大、跨集群多租户环境复杂、血缘链路长，且任一加工环节异常均需人工逆向穿透 3 至 5 层排查的问题开展智能化改造。为此，Eva 构建五层架构，通过意图识别、任务编排、血缘分析、代码与数据检索、长期记忆和知识图谱等技术，自动补全问题上下文并定位异常节点与根因。应用后，约 90% 的排查类问题可由智能体自助回答，其中约 50% 无需人工判断即可直接解决。

（四）数据智能安全与合规

随着企业数据来源、使用主体和流通场景不断增加，敏感数据可能在采集、传输、存储、分析、共享和销毁等环节发生泄露、越权使用或违规流转；应用程序接口、智能体工具调用和跨系统数据交换的普及，又进一步扩大了数据暴露面。传统安全与合规工作主要依赖人工审查、静态权限和预设规则，难以持续盘点海量数据资产、适应法规与业务规则变化，也容易出现审查覆盖不足、告警误报较多和风险处置滞后等问题。针对上述痛点，数据智能体通过融合多模态识别、语义理解、知识图谱、检索增强生成、行为建模和工具调用等技术，

将法规标准、内部制度和安全策略转化为可执行的审查与控制逻辑，形成“资产发现—敏感识别—风险研判—策略执行—处置留痕”的人机协同模式。

当前，数据智能安全与合规已由单一的数据脱敏、权限控制或合规问答工具，逐步扩展为覆盖数据全生命周期和智能体运行全过程的综合能力体系。在数据使用环境中，相关产品通过本地化部署、数据不出域、静态加密、恶意文件检测、代码沙箱、最小权限、结果水印和到期销毁等方式建立安全执行边界；在风险监测环节，通过汇聚数据库日志、应用程序接口日志、系统日志和访问身份信息，构建数据资产与用户行为关联图谱，动态学习正常访问基线，识别异常查询、批量导出、越权调用和数据外传等风险，并还原完整访问链路；在合规审查环节，则将法律法规、行业标准、企业制度和历史案例接入知识库，通过信息提取、规则匹配、分类分级、模型推理和人工复核，完成合同、数据产品、产权登记和流通交易材料的智能审查。同时，部分产品开始面向智能体自身建立资产测绘、动态鉴权、意图识别、安全护栏、红队测试和全链路审计能力，使安全管控由事后检查转向事前预防、事中阻断和事后追溯。

例如，新华网建设的 GEO 智能体平台针对人工智能生成内容存在的语料投毒、虚假信息、深度伪造及传播过程难追溯等问题，依托央媒权威内容资源，构建权威信源库和可信信源分级体系，融合多模态合规检测、内容溯源和价值观对齐等技术，对内容生成到分发全链路追踪。

六、数据智能体的挑战及展望

（一）产业落地核心挑战

1. 模型幻觉与执行效能挑战（“不精确”的瓶颈）

这是数据智能体技术落地的底层短板，叠加多重技术缺陷导致落地受阻、结果失准。大模型内生幻觉易生成虚构数据与错误逻辑，RAG检索精度不足、上下文混淆进一步加剧信息偏差，再加上任务编排能力薄弱，共同构成技术痛点。面对企业高频迭代的数据库 Schema 与定制化技术栈，智能体泛化能力不足，需投入大量人力进行微调适配；跨系统长链路任务中，调度鲁棒性欠缺，易陷入死锁、无限重试等困境；多表深度关联、跨业务线精算等复杂场景下，因果推理断层引发误差累积，最终导致“流程跑得通、结果算不准”，严重制约落地进度与应用价值。

2. 自主行为越权和可控性挑战（“不可控”的黑盒）

聚焦数据智能体自主决策能力背后的不可控风险，其行为易突破预设边界，形成难以管控的“黑盒困境”。自然语言交互的开放性大幅拓宽系统攻击面，恶意用户可通过提示词注入绕过应用层安全防御，诱导智能体执行删库、篡改数据、窃取核心机密等高危操作；跨域协同与联邦查询场景中，智能体易因过度关联分析违背数据最小可用原则，无意中泄露个人敏感信息。此外，多智能体集群规模化部署后，传统静态权限管控体系完全失效，动态鉴权、全流程行为审计机制缺失，智能体操作轨迹难以追溯，进一步放大“控不住”的安全隐患。

3. 人类自主权弱化与权责认定模糊（“不敢用”的红线）

随着数据智能体由辅助分析进一步进入数据处理、业务决策和操

作执行环节，一旦出现错误分析、越权调用、数据泄露或业务损失，责任究竟应由模型与平台提供方、系统建设与运营方、数据提供方、业务使用人员，还是最终审批人员承担，尚缺乏清晰统一的认定规则。特别是在多模型、多工具和多智能体协同场景中，问题可能源于训练数据偏差、模型判断失误、权限配置不当、工具执行异常或人工审核缺位，责任链条较长且难以追溯。同时，过度依赖智能体也可能削弱人员的判断、复核和最终决策权。因此，需要明确“谁开发、谁部署、谁授权、谁使用、谁审批、谁负责”的权责边界，保留重大事项的人工确认与接管机制，并通过全过程日志、决策依据和操作记录实现责任追踪。

（二）技术发展趋势

1. 架构范式演进：从单智能体执行到多智能体协同

数据智能体正加速脱离“前端交互插件”的表层定位，深度内化为企业 IT 架构的核心路由中枢。传统单智能体受限于能力边界，难以覆盖数据全链路复杂需求，而多智能体协同架构将成为主流：聚焦数据采集、治理、分析、合规、安全等垂直领域的专职智能体，将依托 A2A、ACP 等标准化通信协议，构建流水线式的协同作业集群。该集群以全局视角统一调度底层异构数据资源与业务系统，将原本离散的信息孤岛动态编织为语义互通、能力互补、即需即用的智能数据网络，实现复杂数据任务的柔性拆解与高效交付。

2. 内生机制突破：从技能固化到自主进化与持续迭代

当前数据智能体本质上是“预训练知识+人工预设规则”的静态固化执行体，核心能力在模型训练与系统部署阶段即被锁定，无法在运

行中自主学习新业务知识、优化决策逻辑或拓展能力边界。面对业务规则迭代、数据模式变化、新场景涌现等动态需求，只能依赖人工重新标注数据、微调模型或修改配置，存在响应滞后、维护成本高、适配性差等突出问题。未来，数据智能体将构建低人工干预的闭环自主进化机制。通过长时记忆引擎沉淀实战经验，依托强化学习与反馈调优持续迭代认知与决策能力，并反向赋能数据基础设施优化，形成“数据-经验-智能”的正向循环，实现能力的持续自主生长。

3. 应用边界拓展：从数字世界数据智能体到虚实融合的具身数据智能体

数据智能体的作用范围将从纯数字世界延伸至物理世界，催生具备实体感知与执行能力的具身数据智能体。这类智能体通过协议深度连接工业传感器、IoT 设备、机器人、自动驾驶系统等物理终端，不仅能处理数字空间的结构化与非结构化数据，更能实时感知物理世界的状态变化，完成数据采集、实时分析、决策生成到物理动作执行的全链路闭环。在工业制造、智慧物流、城市治理等场景中，具身数据智能体将实现物理世界与数字世界的实时映射与双向联动，推动数据价值从“辅助决策”向“直接驱动物理生产”升级。

（三）行业生态未来前景展望

随着技术迭代与市场需求的驱动，数据智能体已正式跨越技术验证期。展望未来，产业将在能力升级、协同深化与标准完善三大主线上，迎来规模化落地的黄金时代。

1. 范式重构：数据智能体效能跃升，与人工智能发展同频共振

数据智能体的核心能力将持续迭代升级，在多模态数据处理、复

杂任务编排、自主决策推理等方面实现突破，彻底破解当前数据利用不充分、价值转化率低的痛点。通过深度融合数据编织、长时记忆引擎等底层技术，智能体将实现对全域数据的精准感知、高效治理与深度挖掘，推动数据要素从“静态沉淀”向“动态增值”转变，充分释放数据乘数效应。同时，随着智能体与数据资源的协同互促、同频共振，将构建起“数据赋能模型、模型驱动智能体、智能体激活数据”的良性循环，为人工智能产业高质量发展提供更加强劲的核心支撑，推动 AI 从技术突破向规模化价值落地跨越。

2. 壁垒消融：从单体智能体向智能体互联网演进，全链路协同发展

未来，数据智能体将打破单体运行的局限，逐步向多智能体协同、跨域互联的“智能体互联网”演进。随着 MCP、A2A、ANP 等通信协议的广泛应用，不同领域、不同功能的智能体将实现高效互联、能力互补，形成协同联动的智能体集群。产业链上下游将呈现清晰的分工协同格局：底层厂商聚焦算力、基础模型等底座能力优化，中游企业专注于开发框架、工具链与垂直场景解决方案，下游应用方则聚焦场景落地与价值转化，通过全产业链协同发力，降低开发与落地成本，提升产业整体效能，推动数据智能体生态向多元化、规模化方向发展。

3. 合力共建：行业标准逐步完善，产业规范发展

随着数据智能体应用场景的不断拓展与产业规模的持续扩大，行业标准体系将逐步完善，填补当前技术规范、能力评估、安全管控等领域的空白。未来，将逐步形成涵盖智能体技术研发、部署实施、能力评测、安全合规等全流程的标准规范，明确智能体的技术要求、接口规范、权责边界与安全底线。标准的完善将有效解决当前行业内技

术碎片化、能力参差不齐、互操作性差等问题，降低跨主体、跨领域协同的门槛，规避技术与安全风险，引导产业从“野蛮生长”向“规范有序”转变，为数据智能体的规模化落地与长期健康发展筑牢基础。

中国互联网协会人工智能工作委员会