



大数据技术标准推进委员会
Big Data Technology and Standard Committee

数据智能研究报告

(2025 年)

CCSA TC601 大数据技术标准推进委员会

2025年6月



版 权 声 明

本报告版权属于 CCSA TC601 大数据技术标准推进委员会，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：CCSA TC601 大数据技术标准推进委员会”。违反上述声明者，本组织将追究其相关法律责任。

编制说明

本报告的撰写得到了数据智能领域多家企业与专家的支持和帮助，主要参与单位与人员如下。

参编单位：大数据技术标准推进委员会、交通银行股份有限公司、人保信息科技有限公司、南方电网数字平台科技（广东）有限公司、中国联合网络通信集团有限公司、中国联合网络通信有限公司智网创新中心、联通数据智能有限公司、中电信人工智能科技（北京）有限公司、中移动信息技术有限公司、中国民航信息网络股份有限公司、中国联合网络通信有限公司智能城市研究院、中国移动通信有限公司研究院、北银金融科技有限责任公司、中电云计算技术有限公司、北京数巅科技有限公司、亚信科技（中国）有限公司、上海零数众合信息科技有限公司、上海爱数信息技术股份有限公司、江苏钟吾大数据发展集团有限公司、北京谊慧信息技术有限公司、杭州数梦工场科技有限公司、北京盛汉律师事务所、北京交通大学、腾讯云计算（北京）有限责任公司、北京枫清科技有限公司、北京数势云创科技有限公司

参编人员：田稼丰、王卓、王超伦、尹正、杨靖世、张越、姜春宇、马鹏玮、康宸、周一帆、梅宇婷、查毓灵、杨琦、杨光、凌敏、徐乐西、葛迪、蔡志强、丁宏伟、李阳、唐旭、黄启洲、王项男、王伯君、裴亚、袁雪梅、史贇、陈卓、任鹏飞、张若周、周钢、肖巍、张磊、杨健、强薇、董正浩、郭中梅、吴博、金洁华、刘小成、宁艺昭、李亚兰、刘现亮、何昌华、蒋锡茸、张国贤、梅珂夫、王立冬、兰春嘉、杨珍、李基亮、宁静、许凤玲、杜成、刘沂鑫、周丞、王璐、

徐鹏飞、何转琴、郭宁、肖敬仁、姜怀舒、姜雪莹、冀振燕、熊泽昌、
徐圣凯、徐晓敏、田小康、陈海洋、王传阳、孙建新、韩秀锋

前 言

近年来，伴随生成式人工智能技术的突破，人工智能成为数据要素价值释放的关键途径和有效手段。大数据产业与人工智能产业间的关系也从单向赋能转化为深度融合，数据智能产业应运而生。数智协同，共同形成改变物理世界格局、重塑数字世界秩序的核心力量。

2024 年至今，数据智能产业剧烈变革，取得系列突破。**从政策角度来看**，我国密集出台一系列政策，开展“人工智能+”行动，推动高质量数据集建设，加速数据智能技术与各行业的深度融合，为数据智能产业注入了强劲动力。**从技术角度看**，DeepSeek 的横空出世颠覆了“高算力和高投入是发展人工智能唯一途径”的固有认知，引领从业者进入以算法和模型架构优化为主，同时高度重视数据质量与规模、理性提高算力的新时期。**从应用角度看**，智能体的出现进一步降低了数据智能技术的应用门槛，推动了数据智能技术在更多领域的落地，加速了千行百业的数智化转型。

为总结数据智能产业发展现状，研判发展趋势，指引企业顺利实现数智化转型，大数据技术标准推进委员会牵头，联合行业专家和头部企业共同编制《数据智能研究报告（2025 年）》。本报告聚焦数据智能这一话题，系统性厘清数据智能技术体系，深入剖析数据、算法、应用、安全等数据智能具体领域的现状问题，以期为企业未来的数据智能实践提供参考。由于时间仓促，水平所限，本报告仍有不足之处，欢迎联系 tianjiafeng@caict.ac.cn 交流探讨。

目 录

一、数据智能综述.....	8
（一）数据智能的价值与意义	8
（二）国内外数据智能政策综述	9
（三）数据智能产业综述	12
（四）四大领域协同驱动数据智能产业发展	15
二、高质量数据为数据智能实践奠定坚实基础.....	17
（一）高质量数据进入建设元年	17
（二）三大瓶颈限制各行业高质量数据建设	18
（三）高质量数据建设呈现出三大趋势	20
三、高水平算法为数据智能实践提供核心动力.....	23
（一）以大模型为代表的人工智能技术加速渗透	23
（二）智能化水平进一步提升面临三重问题	24
（三）大模型研发应用支撑能力呈三大发展趋势	28
四、高价值应用是数据要素价值释放的关键环节.....	31
（一）数智应用百花齐放，场景渗透持续深化	31
（二）数智应用落地面临三方面挑战	33
（三）AI 驱动数智应用呈现四大趋势	35
五、高安全防护护航数据要素流通与 AI 融合新范式	39
（一）智能化数据安全治理体系加速升级	39
（二）数据智能浪潮中面临三大安全挑战	40
（三）智能驱动的下一代数据安全呈现出两大趋势	42
六、总结和展望.....	44

图 目 录

图 1 全球数据智能企业数量国家分布 12

图 2 2019-2024 全球数据智能领域投融资规模 13

图 3 全球及中美数据智能高水平论文发文量 14

图 4 2014-2024 年各国数据智能专利授权数 15

图 6 《人工智能数据工程能力要求》标准框架 21

图 5 《非结构化数据处理平台技术要求》标准框架 22

图 7 以大模型为代表的人工智能技术体系概览 24

图 8 《智能中台能力成熟度模型》框架 29

图 9 数智应用体系概览 31

图 10 《企业知识工程能力成熟度模型》标准框架 36

图 11 《智能体研发运营(AgentOps)能力成熟度模型》标准框架 37

图 12 《大模型驱动的数字员工平台技术要求》标准框架 38

表 目 录

表 1 各行业数智应用落地的头部场景 32

一、 数据智能综述

（一） 数据智能的价值与意义

近年来，智能领域迎来“量变引发质变”的临界点，相关技术、产业进入剧烈变革期。自 1956 年人工智能（AI）概念诞生以来，智能计算领域历经多个阶段的技术方向探索，逐渐收敛在深度学习这一主线，但仍以“决策式人工智能”为主要发展领域。近两年，在以 Transformer 模型为代表的算法、极致算力支撑下的千亿级模型参数、大规模高质量数据集三者共同的作用下，生成式大语言模型的应用效果出现跨越式提升。以 GPT、DeepSeek 为代表的大模型能实时对图像、音频、视频等多种形式输入进行理解，根据要求完成高效问答、内容生成等多种任务，由此“生成式人工智能”及相关应用的发展成为全球焦点，带动人工智能技术产业进入剧烈变革期。

伴随智能领域变革，“数据”与“智能”间的发展关系亦呈现两点重要变化，一是“智能”将成为“数据”价值释放的主要路径，数据将成为智能成效进一步跃迁的胜负手，二是智能化技术开始反向助力数据技术发展和非结构化数据应用。数据与智能正加速融合，由此“数据智能”的概念也应运而生。数据智能的概念可以初步概括为，以全形态数据为关键资源，以大数据和人工智能深度融合后的新技术体系为关键手段，以决策式、生成式人工智能和传统数据应用形式协同应用于生产生活各领域为最终目标，由此形成的新兴生产生活方式，以及相应延展出的新技术、新产业、新生态。

数据智能不仅提升了决策效率和资源配置能力，还推动了创新和用户体验的提升。数据智能技术能够将海量数据转化为具有深度洞察力的知识和决策依据，提升决策效率、优化资源配置，为各行业带来前所未有的价值和变革，成为推动全球数字化转型的核心引擎。数据智能将逐渐为企业、产业、社会等各层面带来新的价值和意义。

在企业层面，数据智能的实践能提升企业从数据中提取有效信息、精炼转化为知识、最终指导决策这一过程的总体效率。**在产业层面**，数据智能的实践在直接带动相关技术服务产业发展的同时，还将带来模式创新和对生产关系的重塑，改善产业链总体产出效率。**在社会层面**，数据智能的实践能直接提升信息、知识在全社会范围内的流动效率，同时借由对信息的互通和技术的应用强化总体协同性，优化社会资源的配置效率。

（二） 国内外数据智能政策综述

数据智能的快速发展促使全球多国综合考虑发展与监管，加快相关政策出台。

美国出台法案推动人工智能研发创新和应用落地。2023 年 5 月，白宫更新了《国家人工智能研发战略计划》，旨在加大联邦政府在人工智能研发方面的投资，确保美国在开发和使用可信赖人工智能系统方面继续保持全球领先地位，并为未来的人工智能研发做好准备。2024 年 3 月，美国白宫管理和预算办公室（OMB）发布《推进联邦机构使用人工智能的治理、创新和风险管理》指导意见，旨在规范联

邦机构在使用人工智能时的治理、创新和风险管理，确保 AI 使用的合规性和风险管理。

欧洲建立人工智能监管框架规范 AI 的开发和利用。欧盟《人工智能法案》于 2024 年 8 月 1 日正式生效，是全球首部全面监管人工智能的法律，将人工智能系统划分为“不可接受风险”、“高风险”、“有限风险”和“低风险”四类，并对高风险 AI 系统提出包括风险缓解措施、数据集质量、用户信息透明度和人工监督等在内的严格要求。2025 年 3 月，英国正式提出《人工智能（监管）法案》，主要内容包括设立人工智能管理局、明确监管原则、推动监管沙盒计划、规定透明度与知识产权义务等，希望在促进创新的同时，确保人工智能技术的安全性、可靠性、公平性与可追责性。

日韩通过制定战略规划为 AI 产业提供政策支持和法律保障，鼓励技术应用、创新和产业发展。韩国制定《人工智能基本法》，通过设立国家人工智能委员会和人工智能安全研究所、为 AI 研发、科研数据集构建及人工智能中心建设提供支持、明确处罚规则等手段构建一个系统化的 AI 监管框架，同时为 AI 产业提供政策支持和法律保障。日本出台《人工智能运营商指南》对人工智能开发人员、供应商和业务用户提出安全要求，包括但不限于确保人工智能系统的透明性和可解释性、提升数据质量、加强隐私保护以及建立有效的风险管理系统，降低数据隐私、偏见、安全性等风险。

我国发布一系列政策文件，为数据智能产业发展注入强劲动力。国家层面，2023 年 8 月施行的《生成式人工智能服务管理暂行办法》

及 2024 年 9 月发布的《人工智能生成合成内容标识办法（征求意见稿）》等法规，明确了生成式人工智能“提供者”在内容生产、数据保护、隐私安全等方面的法定责任及法律依据，要求网络信息服务提供者对 AI 生成的合成内容进行标识，为人工智能技术的规范发展提供了明确的法律依据。2023 年 12 月发布的《“数据要素×”三年行动计划（2024—2026 年）（征求意见稿）》，部署了推动高质量数据集建设、支持大模型训练与应用、优化数据流通环境、推动数据要素与行业结合等一系列行动，推动人工智能技术与各行业的深度融合，为数字经济的高质量发展提供有力支撑。2025 年 5 月，国家数据局发布《数字中国建设 2025 年行动方案》，方案部署了“人工智能+”、数据产业培养等八项重点行动，旨在通过技术创新和应用场景拓展，赋能千行百业，加速经济社会的数字化转型和高质量发展，推动各行业的智能化升级。**地方层面**，北京、上海、深圳、江苏、山东等省市纷纷出台人工智能、大模型、数字经济相关政策，通过建设智算中心、打造示范性应用项目、设置人工智能创新发展试验区等方式，推动数据智能在不同领域的应用和发展，为数据智能产业的健康发展提供了有力的政策支持，突显出各地在数据智能领域布局的积极性。

（三）数据智能产业综述

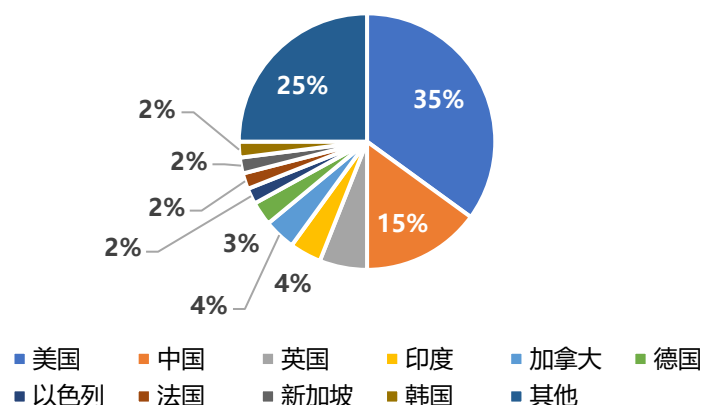


图 1 全球数据智能企业数量国家分布

企业层面，全球数据智能产业相关企业加速涌现，企业总量超三万家。据中国信通院统计，截至 2024 年 12 月，全球共有数据智能企业 31490 家，其中，美国企业达 11090 家，占全球总数的 35%，中国企业 4696 家，占全球总数的 15%。从地域看，我国数据智能企业主要集中在经济发达地区。北京、广东、上海、浙江四省市数据智能企业数量合计占全国数据智能企业数量的 76%，东部 10 省市数据智能企业数量超 4000 家，约占全国数据智能企业的 87%。未来，随着大模型技术在各行业领域的落地应用，数据智能企业数量将持续增长，支撑产业规模持续扩大。

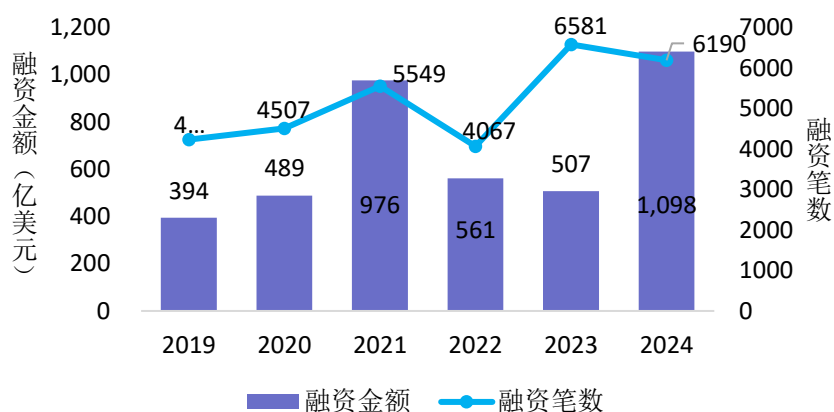


图 2 2019-2024 全球数据智能领域投融资规模

投融资层面，数据智能企业投融资活跃，融资金额创历史新高。

最近一年，随着大模型技术应用快速发展，数据智能产业受到资本持续关注，融资金额创历史新高。2024 年全球数据智能融资金额达 1098 亿美元，同比增长达 117%，融资笔数达 6190 笔，同比下降 6%。全球数据智能融资占全行业融资比例再次上升并创下新高。数据智能融资占全行业融资比例从 2022 年的 6.1% 上升至 2024 年的 13.6%，并在 2025 年 Q1 进一步升至 24.4%。资本市场对于数据智能产业的良好预期，对于产业新增长点的投资带动作用非常显著，月之暗面、智谱 AI、xAI、Athropic 等为代表的一批数据智能领域企业在 2024 年均完成大笔融资。面对数据智能应用在资本市场和用户市场的火热态势，国内外科技巨头也纷纷入场，国外以微软、谷歌、Meta 等为代表，国内以阿里、字节跳动等为代表，一方面积极研发企业级数据智能相关产品，另一方面以开放接口等方式与其他企业合作开发应用，加速构建数据智能产业生态。

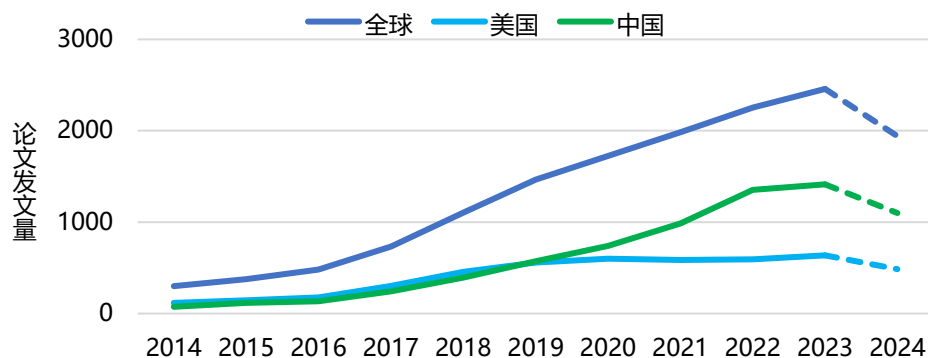


图 3 全球及中美数据智能高水平论文发文量

科研层面，全球数据智能领域高水平论文发文量逐年增加，中美专利授权量领跑全球。论文方面，2014-2024 年，全球数据智能领域高水平论文发文量逐年增加，总量超 14000 篇，其中 2024 年全球数据智能领域论文发布量超 25 万篇，高水平论文达 1941 篇。中国高水平论文自 2019 年超过美国后逐年上升，2024 年发文量达 1100 篇，美国为 486 篇。10 年来，中国论文发文量超 51 万篇，美国发文量超 30 万篇，印度、英国、德国分列第三、第四、第五位，中国整体发文量高于美国，但美国企业科研更为活跃，科技巨头微软、谷歌、IBM 分别发文 5971 篇、5363 篇、3786 篇。专利方面，全球数据智能专利授权比例显著提升，2019-2021 年，全球数据智能专利授权量占申请量的比例不足 30%，2022-2024 年，这一比例提升至 58%。从国家来看，全球数据智能专利申请和授权量集中在中美两国，10 年来，中国专利授权量超 35 万，美国授权量超 14 万，其他国家专利数量与中美差距较大，韩国、日本、德国分列第三、第四、第五位。

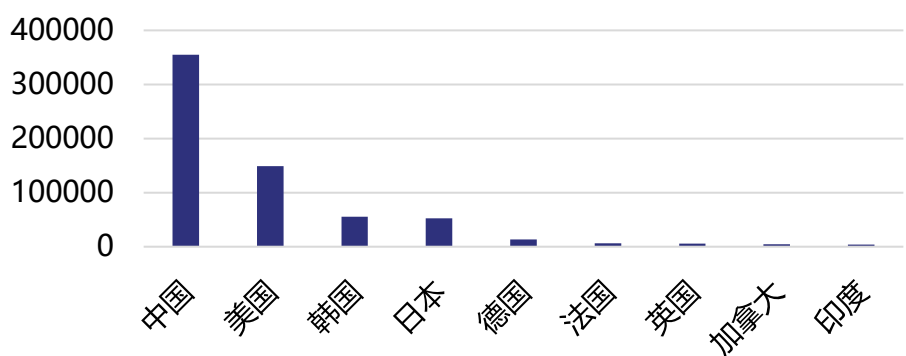


图 4 2014-2024 年各国数据智能专利授权数

人才层面，美中两国成为全球数据智能人才聚集高地，人才培养体系正加速构建。随着各国在数据智能领域竞争的日益激烈，人才正成为推动产业生态发展、抢占国际领先地位的核心资源，不管是企业端还是高校机构，对数据智能人才的重视度都在不断增加。据美国保尔森基金会发布的报告显示，中国和美国是顶级数据智能人才的主要来源地和目标工作地，70%的顶级数据智能人才在中国或美国的机构中工作，65%的顶级数据智能人才出自中美两国。中国是全球最大的顶级数据智能人才输出国，在中国接受本科教育的顶级（前 20%）数据智能人才占全球 47%，GPT-4 团队的核心贡献者名单中，约 20%的研究人员来自中国。

（四） 四大领域协同驱动数据智能产业发展

随着产业和技术的持续发展，数据智能产业内涵逐步清晰，包括数据、算法、应用、安全四大核心领域。四大领域相辅相成、紧密协作，构筑了数据智能产业的完整链条。

数据是数据智能产业的基石。高质量的数据能够帮助模型更准确地学习和预测，决定了模型的精度与应用的可靠性，为数据智能实践

奠定坚实基础。

算法是数据智能产业的引擎。从机器学习到深度学习，从强化学习到生成式大模型，算法的每一次突破都为数据智能带来了新的可能性。高效的算法能够更好地挖掘数据中的价值，提升模型的效率和准确性，为数据智能实践提供核心动力。

应用是数据智能产业的牵引。应用是数据智能产业从技术理论形成生产力的关键。将数据智能应用到企业运营效率提升、用户体验优化、业务模式创新后，数据智能产业才能持续获得投入，才具备长期发展的生命力。

安全是数据智能产业的保障。随着数据智能应用的广泛普及，数据泄露和针对模型的恶意攻击的风险也随之增加。完善的安全技术能够有效抵御这些威胁，防止数据被非法获取、篡改或滥用，为数据与AI融合新范式保驾护航。

二、 高质量数据为数据智能实践奠定坚实基础

（一） 高质量数据进入建设元年

高质量、多维度的数据集是支撑大模型训练、应用的关键基础。近年来，随着数智应用的进一步发展，为进一步提升模型准确性及开发效率，高质量数据集的建设受到广泛关注。高质量数据集**一是**能为机器学习和人工智能模型提供可靠的基础，显著提升模型的准确性、稳定性和泛化能力，使其在复杂多变的现实场景中更好地发挥作用；**二是**有助于减少偏差和不公平现象，确保模型的决策过程符合伦理和法律要求，增强公众对技术的信任；**三是**优化数据管理和分析流程，提高研究和开发效率，降低因数据问题导致的资源浪费。

国家高度重视高质量数据集建设工作，出台诸多政策推动高质量数据集建设。2024 年 12 月，国家发改委等部门联合印发《关于促进数据产业高质量发展的指导意见》，首次提出“高质量数据集”，支持企业开发高质量数据集。同月，《关于促进数据标注产业高质量发展的实施意见》进一步指出加强重点行业领域数据标注，建设行业高质量数据集。2025 年 2 月，高质量数据集建设工作启动会在北京召开，国家发展改革委、教育部、科技部、工业和信息化部等 27 个部门参会，下一步，高质量数据集建设工作将积极推进落实“人工智能+”行动，加快推动形成一批标志性成果，赋能行业高质量发展。2025 年 5 月，《数字中国建设 2025 年行动方案》指出要深度挖掘人工智能应用场景，积极开展人工智能高质量数据集建设，加强交通、医疗、金

融、制造、农业等重点领域数据标注，建设行业高质量数据集。

我国各行业企业已开展高质量数据集建设实践，取得显著成果。在第八届数字中国建设峰会期间，国务院国资委集中发布了首批 10 余个行业 30 项央企人工智能行业高质量数据集优秀建设成果，首批成立交通物流、绿色低碳、智慧能源等三大行业中央企业数据产业共同体。2024 年 5 月，成都、沈阳、合肥、长沙、海口、保定、大同 7 个城市入选国家数据局首批承担数据标注基地建设任务的城市名单。此外，湖北、天津、杭州等地市也已出台政策和鼓励措施，并开始围绕工业、金融、农业等行业征集高质量数据集。

（二） 三大瓶颈限制各行业高质量数据建设

当前，高质量数据建设过程中面临缺乏落地实践方法论、技术能力存在短板及数据流通机制不完善三大瓶颈，严重制约了数据智能技术在垂直领域的深度应用与价值释放。

1. 缺乏落地实践方法论

当前，业界在高质量数据集建设方面面临实践方法论缺失问题。**一方面**，行业大模型的数据需求呈现出极强的多样性，不同行业、部门对模型场景数据的需求千差万别，涵盖了从分析、决策到生成等多种任务需求，同时涉及预训练、微调、反馈强化学习等不同阶段。这一复杂过程需要对来自多个数据源、多种数据类型的数据融合对齐，极大地增加了数据处理和治理的复杂度，而目前在这一领域缺乏成熟的实践方法论进行有效指导。**另一方面**，构建大模型数据集涉及数据

采集、数据清洗、数据标注、质量评估等核心环节，每个环节都需要根据大模型数据集的规模、类型、行业垂直属性等特征进行针对性的技术研发和适配。然而，由于业界缺乏高质量数据集构建的系统性方法论，导致面向大模型数据集的治理方法和经验严重不足，难以满足数据智能产业快速发展的需求。

2. 技术能力存在短板

数据集的构建并非简单的数据堆积，而是涉及数据采集、清洗、处理、质量评估等一系列复杂流程。**在数据采集阶段**，不同系统、应用和平台产生的数据格式各异，导致数据的兼容性和一致性问题突出，例如，图像数据可能来自不同的设备，其分辨率、色彩空间等参数不一致，文本数据可能来自网页、文档等多种形式，格式混乱，降低数据采集的整体效率；**在数据清洗阶段**，数据中可能存在缺失值、噪声数据、重复数据等问题，这些问题会影响模型的训练效果，需要进行数据清洗。然而，现有的清洗工具可能无法完全自动化地处理这些问题，仍需大量人工干预；**在数据处理阶段**，多模态数据处理技术不足，缺乏针对多模态数据的统一融合算法，针对半结构化和非结构化数据的处理的效率较低，难以满足上层模型的训练需求；**在质量评估阶段**，现有的数据质量评估工具大多面向通用数据，对特定领域或特定类型的数据支持不足，例如，对于时空数据、多模态数据等，现有的评估工具可能无法准确评估其质量。

3. 数据流通机制不完善

在基础设施层面，目前我国正积极建设数据流通利用基础设施，打造企业、行业等可信数据空间，技术已基本可实现数据安全、隐私保护等能力，但整体机制建设尚未落地，多方主体流转过程中的权益保障、问题追责等难点仍未解决，导致高价值数据仍更多在互信主体间流通使用。在企业数据层面，行业高价值数据多掌握在平台型企业或产业龙头企业中，出于数据安全、商业竞争、责任规避等原因，数据资产往往被封闭运营，极少以结构化、标准化形式向外部提供，而一些尝试对外提供数据资源的企业也普遍面临定价机制不清、数据使用授权不规范、数据变现路径模糊等障碍，进一步削弱数据流通积极性。在当前政策积极引导的背景下，企业应尽快补齐机制短板，加强数据流通底座和相关技术工具研发，推动数据可信汇聚与使用。

（三） 高质量数据建设呈现出三大趋势

随着数字经济步入深水区，数据资源从“工具”到“资产”乃至国家战略资源，高质量数据的重要性日益显现。面对日益丰富的海量数据和复杂的应用场景，仅仅追求数据的数量与规模已经无法满足智能化决策和业务精细化管理的需求，提升数据质量、数据价值已成为企业和政府部门工作的重心。当前和未来一段时间，高质量数据建设呈现三大特点：

1. 数据集建设运营方法论逐步成熟

随着数智技术的飞速发展，高质量数据集的建设与运营方法论正逐步走向成熟，成为推动数据智能应用落地的关键支撑。在建设阶段，

企业必须从战略高度出发，明确数据集的业务目标和应用场景，确保其构建与企业的核心需求紧密相连。在运营阶段，企业需建立数据更新与维护机制，通过制定量化指标和定期评估机制，实时监控数据集的质量状况，并及时进行优化调整。随着这些方法论的逐步成熟，高质量数据集的建设与运营将更加系统化、规范化和高效化，为模型训练及数据智能应用开发提供坚实基础。为助力各方开展高质量数据集的建设运营工作，中国信通院推出《高质量数据集建设运营能力成熟度模型》《人工智能数据工程能力要求》等多项标准，为企业高质量数据集建设与运营能力提升明确路径与目标，推动企业数智化转型。

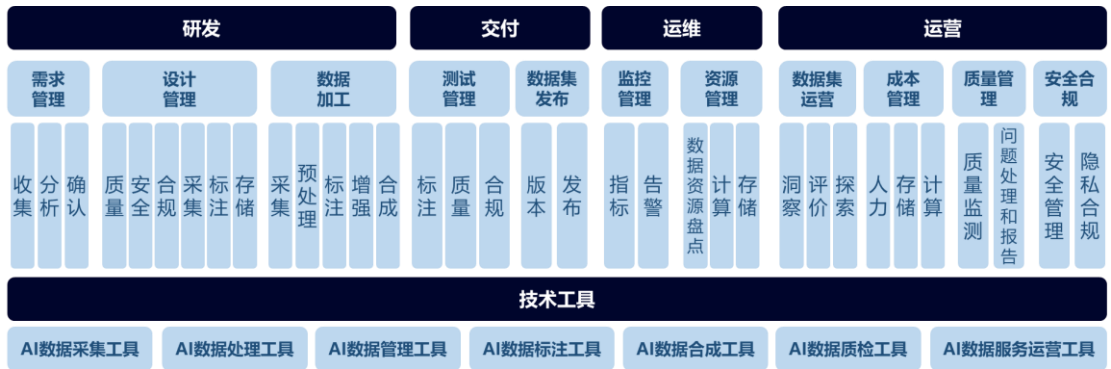


图 5《人工智能数据工程能力要求》标准框架

2. 多模态数据处理技术加速创新

多模态数据处理技术正迎来快速发展的关键时期，一方面，多模态数据处理技术工具进一步完善。得益于数据智能技术的持续发展，智能化能力的引入提高了多模态数据处理的效率和质量，为高质量数据集建设奠定坚实基础。例如，在数据清洗环节，借助先进的大模型或机器学习算法，工具能够自动检测并修复缺失值、去除噪声数据、识别重复记录，极大地减少了人工干预，显著提升数据的可用性。另

一方面，融合深度与广度将不断拓展，从图像与文本结合，逐步迈向语音、视频、传感器数据等多源数据的深度融合，实现更全面、立体的信息感知与理解，为复杂场景提供更精准的决策支持。例如，在智能驾驶领域，多模态技术将助力车辆更精准地感知路况并做出快速反应。

非结构化数据处理平台技术要求						
数据接入	数据存储	数据处理		数据管理	数据服务	安全保障
多模态数据接入	多模态数据存储	统一计算框架	图像处理	多模态元数据采集	数据检索	权限管控
多渠道数据接入	分布式存储	分布式计算	音视频处理	元数据管理	在线预览	数据加密
实时离线模式	数据分层存储	workflow编排	文本处理	数据质量稽核	数据开放服务	安全审计
在线接入	数据向量化存储	数据预处理	资源调度	数据标签	分布式索引	服务监控
本地接入	多协议接口访问	数据过滤	数据增强	数据标注	数据水印标识	数据脱敏
接入任务管理	缓存加速	数据清洗		数据标准	多模态关联分析	数据备份与恢复
断点续传	统一命名空间	数据标准化		数据生命周期管理		数据分类分级
	高可用与弹性扩展存储架构	处理任务管理		数据质量报告		
	数据存储压缩			数据价值评估		

图 6《非结构化数据处理平台技术要求》标准框架

3. 多方参与形成协同生态

高质量数据集建设过程中，往往涉及与外部的交易、交互，如数据采买、数据标注、质量评估、利用外部工具进行数据开发等。未来，随着以可信数据空间为代表的数字基础设施的落地，将推动形成高质量数据集协同生态。一方面，协同生态将吸引更多数据提供方与服务方加入，促使更多潜在的数据源被挖掘和利用。在这一过程中不仅能推动高质量数据集与各行业的深度融合，还能将数据开发、质量评估等关键环节交由专业服务方承接，从而显著加速高质量数据集的构建进程。另一方面，随着数据定价、数据权属、收益分配等机制的完善和生态内共识规则的明确。企业将获得更清晰的运营指引和更有力的激励机制，将推动企业高质量数据集持续运营，从而更好赋能数据要素市场发展。

三、 高水平算法为数据智能实践提供核心动力

智能指个体或系统通过感知、学习、推理、决策和适应环境，以实现特定目标或解决问题的综合认知能力。在数据智能的实践中，人们可以通过将智能化技术与生产生活场景充分结合，利用算法模型充分发挥数据资源价值，并以此提升生产生活中各项核心活动的执行效率。在这一过程中，高水平的人工智能（AI）算法模型在其中扮演了重要的角色。

（一） 以大模型为代表的人工智能技术加速渗透

自人工智能（AI）的概念于 1956 年诞生以来，人类致力于创造出具备近似或超越人类智能水平的智能计算系统，协助自身更加高效的解决问题和完成工作。对于人工智能的美好愿景，持续推动着机器学习等一系列相关学科与技术的诞生与发展，伴随着互联网、大数据等信息技术的陆续诞生与普及，通过算法对数据进行分析挖掘以支持决策的应用范式逐渐渗透至各行各业，不断提升人们的日常生活质量及生产效率。近年，生成式大语言模型的应用效果显著，进一步加快了人工智能技术渗透率的提升。

自 2022 年对话式语言生成模型 ChatGPT 发布以来，生成式大语言模型跨越式的应用效果提升再度唤醒人工智能产业的热度，以 Transformer 为代表的算法模型在其中起到了至关重要的作用。随后，在近两年的持续迭代中，强化学习、混合专家模型（Mixture of Experts, MoE）、蒸馏、思维链（Chain of Thought, CoT）、低秩适应（Low-Rank

Adaptation, LoRA) 微调、检索增强生成 (Retrieval-Augmented Generation, RAG) 等技术不断注入,从准确率、可解释性、推理性能、模型性价比等维度提升大模型的各方面能力,诞生出以 GPT-4o、DeepSeek-R1 等为代表的最新成果。

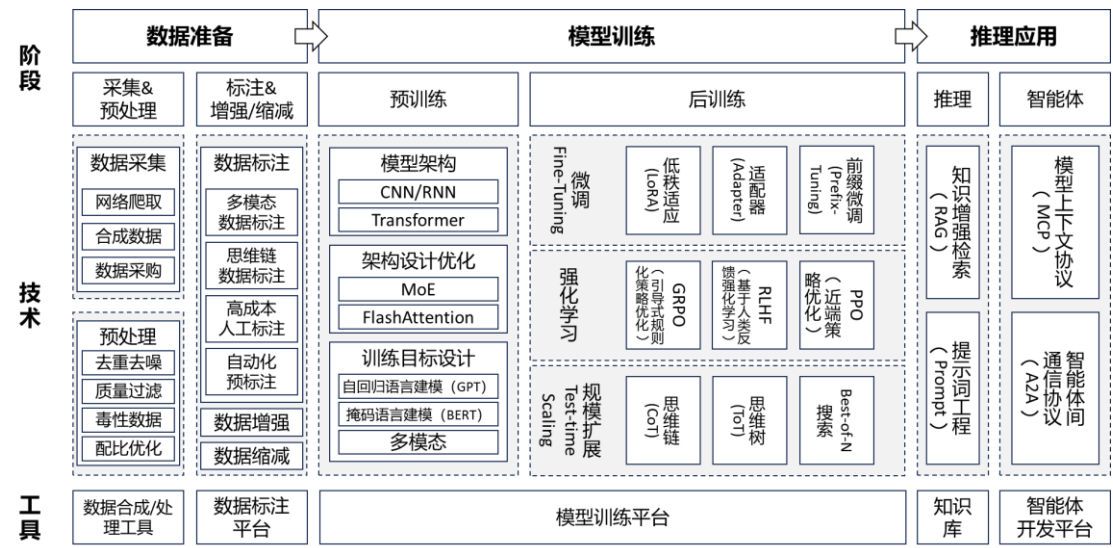


图 7 以大模型为代表的人工智能技术体系概览

模型效果的持续提升,带动以大模型为代表的人工智能技术加速渗透应用至人们的日常生活以及各行各业的实际生产经营中。IDC 数据显示,2024 年中国生成式 AI 占 AI 市场投资总规模的 18.9%,预计 2028 年生成式 AI 投资占比将达到 30.6%,投资规模超 300 亿美元,五年复合增长率达 51.5%。此外,大模型在制造业、医疗、金融等领域的渗透率将持续加深,更多的垂直行业专用模型逐步诞生。同步于行业模型,智能体 (Agent) 这一便于模型更加便捷的执行完成多样化任务的应用形式也成为当下一大发展方向。

(二) 智能化水平进一步提升面临三重问题

当前人工智能技术产业发展正值高潮,但在模型效果、平台工具、

应用形式三个层面仍分别存在待解决的相关问题。

1. 模型效果仍有提升空间

通用大模型的效果虽然还在逐步提升，但面向实际的应用需求仍存在多方面不足。总体来看，在模型算法的效果层面主要存在多模态能力有限、专业领域能力仍有不足、内容安全存在隐患三方面问题：**一是多模态能力有限**，虽然以 GPT-4o 为代表的头部大模型已具备多模态能力，但实际应用中模型在跨模态交互时仍旧存在不同程度幻觉或识别错误的情况，在涉及到众多场景实际需求的空间任务中，仍因空间推理能力不足而导致大量错误。效果不佳与训练、推理算力消耗带来的高成本相结合，综合导致当前模型多模态能力的实际适用性仍十分有限。**二是专业领域能力仍有不足**，在企业的实际生产经营中，仍有大量的应用场景涉及高度专业的行业或领域知识，当前通用模型难以支持。与此同时部分垂直领域专用模型的研发受领域、场景数据供给不足等问题的限制，模型效果仍旧存在缺陷，对于众多高度追求准确性的行业和业务场景来说难以将其投入生产环境。**三是内容安全存在隐患**，大模型迭代演进迅速，但在其效果不断提升的过程中内容安全防护手段仍有不足。当前基于规则、关键词、语义检测的过滤机制，在应对基本的输出内容审核需求之余，面对恶意提示词引导、模型价值观偏离、重要事实性错误等问题作用依旧有限。与此同时，用户问答数据等个人隐私数据，在当前部分大模型数据安全技术能力有限的情况下，仍存在外泄风险。而模型本身训练数据来源是否合规以

及是否存在版权风险等问题，也将为大模型未来的持续发展带来潜在隐患。

2. 技术栈复杂提高使用门槛

以大模型为代表的最新人工智能算法模型，其研发应用过程通常需要一系列较为复杂的工具链支撑，高度专业化的技术工具无形中增加了其技术门槛。当前模型的研发应用过程是包含数据集的标注、预处理、特征提取，模型的设计、预训练、评估、后训练，模型部署、推理应用、监控运营等环节在内的一整套流程。为确保模型研发工作的敏捷性与迭代的便利性，研发流程中的各环节均发展出了一系列高度专业化的程序框架或工具，以便专业开发人员能够高效完成模型研发及迭代优化工作。而在大模型开始加速推广渗透的当下，更多企业试图引入相关技术工具自主进行模型研发、后训练、部署、推理应用，但繁杂琐碎的工具链为其技术选型和采购带来极高的技术门槛和学习成本。

为降低采购及应用门槛，集成化平台工具产品逐步诞生，但实际问题仍难以根除。部分产品供应商逐步推出模型即服务（Model as a Service, MaaS）平台、训推一体化平台等各类平台工具产品，意在向用户提供一套集成模型研发技术栈中复数工具能力的一站式平台化产品，但由于相关概念定义不明确、市场中各企业产品能力不统一、用户已完成部分工具建设等原因，实际项目中仍会出现产品能力缺失导致重复采购、用户存量系统并未利旧导致成本浪费、成套产品采购

导致能力超需求冗余等问题。

3. 应用方式仍在探索尚未形成范式

随着通用大模型的效果提升速度逐步放缓，如何将已有的算法模型充分利用成为产业界更加关心的重要问题。从大模型的应用现状来看，当前仍存在企业端应用场景及形态零散、专业领域应用难度大、智能体应用形式仍需探索等问题：**一是企业端应用场景及形态零散**，产业界近年来对于大模型企业端应用的探索呈现出高度的离散化，较为突出的应用场景如知识检索、数据分析、智能编码、智能运维等场景之间相互独立，包括对话问答、嵌入式插件、无感内置等在内的应用形式形态各异。高度相异的场景特点和应用形式使得大模型企业端应用的渗透率提升难以加速。**二是专业领域应用难度大**，部分行业的生产经营场景高度依赖专业知识，利用大模型时需对通用基础模型进行大量后训练，形成专业领域模型，这一过程不但存在诸多前置条件以及额外成本，同时还存在训练成果弱于传统小模型的风险。此外，将某一特定专业领域模型迁移应用至其他领域的技术难度极高，对大量专业领域逐一训练专业模型的成本又令人难以接受。**三是智能体应用形式仍需探索**，能够依据感知及推理决策执行具体任务的智能体被视为未来大模型应用于企业端的主要形态。但当前智能体的构建多依赖于针对特定任务目标的离散化建设，单一智能体缺乏应对复杂任务和多任务的能力。与此同时，智能体实现特定任务的效果也依赖于决策模型的调用规划以及被调用任务工具的支持水平，导致不同智能体

的应用效果之间存在较大差别。

（三） 大模型研发应用支撑能力呈三大发展趋势

1. 多模态能力持续提升助力专业领域模型发展

多数行业的实际业务场景中，都会涉及大量对于包含图像、视频、音频等在内多种模态对象的理解、推理需求，更好的跨模态推理能力也成为通用基础模型的重点发展方向。当前，国外的 OpenAI、Meta、谷歌，国内的字节、阿里、百度等等大模型头部公司均已发布各自的多模态大模型，后续也将持续提升模型的跨模态理解推理能力。实际上随着通用预训练大模型的效果难以进一步得到飞跃式提升，大模型领域的研究已进入后训练时代，在规模扩展（Scaling）、强化学习、微调（Fine-Tuning）这三类主要后训练技术的支持下，通过针对特定目标模态进行后训练，以此将通用大模型能力迁移扩展至多模态领域带来的成效，相较通用模型自身能力的优化更加显著。同时以 Llama 4 为代表的更多主流大模型开始采用 MoE 架构，多专家领域划分的网络结构使模型对多种模态的支持效率与效果能够得到进一步提升。

2. 智能中台兴起模块化解构模型研发技术栈

大模型的研发应用过程中涉及的技术栈复杂、工具链繁多，以能力建设为主导的中台理念，相较纯粹由技术平台形态牵引的产品概念，更适用于应用方完善自身模型研发应用能力。智能（AI）中台的概念源于数据中台所带动的“中台”概念扩散。智能中台可以理解为企业内部支撑算法模型研发、应用、运营所需的公共能力集合，包含了企

业利用人工智能技术赋能自身智能化转型所需的主要基础支撑性能力。通过能力集合的视角，应用方可对自身数据、算法、算力各方面的支撑情况，即包含数据集管理、算力资源调度等周边支撑能力在内的算法模型加工利用全流程支持情况，进行逐一模块化对标，合理规划自身所需额外建设的能力，简化选型难度，有效进行系统利旧，避免重复建设带来的额外成本。近年来，智能中台的概念逐渐在各行业领域得到实践，中国电信、南方电网等各行业头部企业纷纷推进自身智能中台建设。

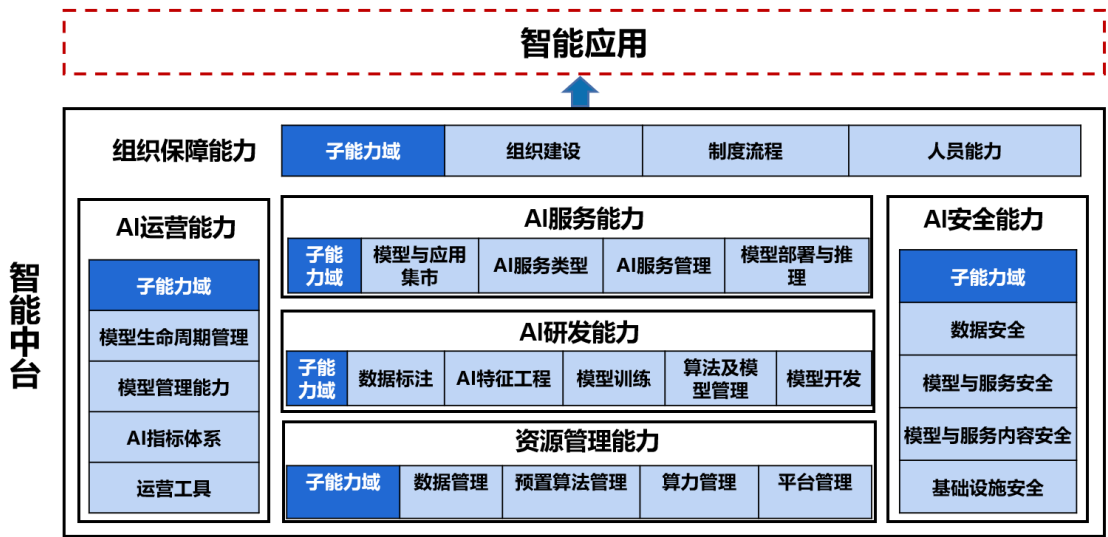


图 8 《智能中台能力成熟度模型》框架

3. 多任务智能体及多智能体协同有望持续突破

模型上下文协议（Model Context Protocol, MCP）逐渐普及，智能体能力上限有望得到进一步拓展。近期，MCP 逐渐流行于越来越多主流大模型的最新版本中，相较于让模型通过传统 Function Call 形式调用外部特定功能，MCP 能够通过更多标准化的预建程序让模型实现更灵活的外部功能集成调用。基于 MCP 的逐渐普及，未来单个智

能体可以更加便利的集成更多外部功能，实现对于囊括多个子任务的复杂任务的有效支持，充分利用模型推理能力，使智能体拥有更加强大的任务规划和丰富的多任务执行能力。与此同时，多智能体协同架构的逐步完善有望弥补单一智能体难以泛化的能力瓶颈。在不断通过技术手段提升单一智能体能力的同时，多智能体协同技术方案也逐步受到关注。今年 3 月发布的 Manus 便采用了多智能体系统架构，对复杂任务能够先进行任务分解，然后由多个不同智能体协同完成各个子任务，最终实现整体任务目标。而在今年 4 月，谷歌发布了用于智能体间通信的开源协议 Agent2Agent（A2A），在 MCP 的基础上进一步允许其他智能体也成为智能体调用的对象，以此加强智能体之间的互操作性和协作能力，使多智能体间协同更加灵活便捷，完成复杂任务的策略和方式更加丰富多变。

在协议层的进展之外，随着智能体的应用加深，AgentOps 作为针对智能体本身的研究运营方法论也更加受到关注。随着企业内部智能体部署规模的增长，对于智能体研发、部署、监控、优化和治理等全生命周期管理的系统性需求也在同步激增。智能体研发运营（AgentOps）是以智能体（AI Agent）为核心对象，强调将传统模型开发与智能体特有的规划、推理、执行、反馈能力相结合，形成端到端的研发运营管理闭环。未来，企业将参照这一方法论，通过构建标准化流程、自动化工具链及跨职能协作机制，提升智能体开发效率，增强智能体运营稳定性，优化多智能体协同能力，促进智能体系统同业务场景的紧密耦合，最终推动智能体的规模化落地与价值释放。

四、 高价值应用是数据要素价值释放的关键环节

数智技术的应用落地包括从数据中提取信息和知识，构筑智能算法和模型，推动决策和行动的整个过程，以实现提高效率、增强体验、驱动创新等目标。对企业来说，数据智能应用是数智化转型的核心组成部分，是释放数据价值的最终一环，也直接决定了数据智能相关实践的最终成效。国家高度出台了一系列政策推动“数据要素×”，“人工智能+”的发展，促进数据要素资源与人工智能技术的应用落地。

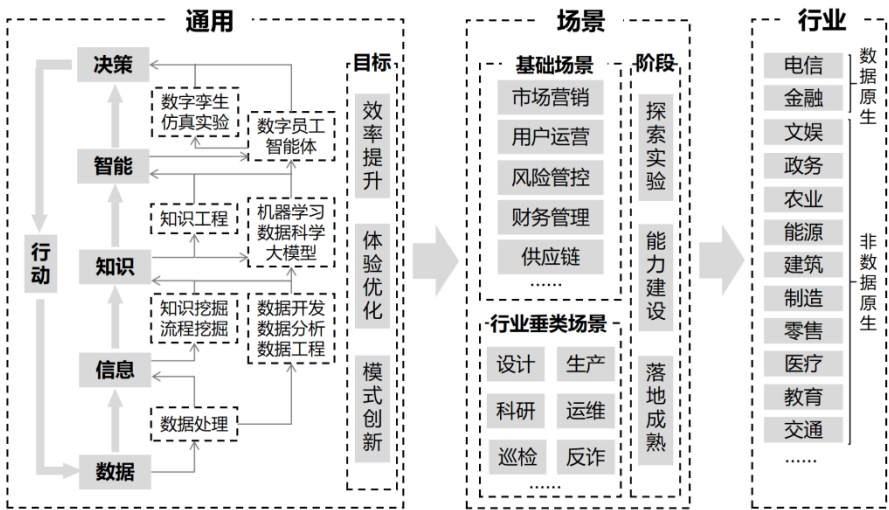


图9 数智应用体系概览

（一） 数智应用百花齐放，场景渗透持续深化

通用范畴来看，数智应用的目标呈现出多样化发展的态势。数智应用的目标可分为效率提升、体验优化和模式创新三类：一是效率提升，即通过数智技术改进原有场景下的工作流程，减少重复和冗余的人工操作；二是体验优化，即通过数智技术对现有业务场景中的痛点进行改造，以提升业务场景中各类参与者的体验；三是模式创新，随着大模型技术的快速落地，数字员工、智能客服、个性化营销、智能

化医疗诊断等新业态、新模式成为发展的重点。大模型的应用落地正在深度改变现有的应用范式，并以智能体为核心落地形态，催生出新的产品和服务，数智应用的目标也更具多元化。

场景范畴来看,大部分业务场景仍处于数智能力建设阶段,营销、运营、管理类场景正在向落地成熟阶段转变。数智技术应用落地较为优先的场景具备高价值、共性强、数据密集、成效明确的特点，一般集中在营销、运营、供应链管理、财务管理、智能风控、智慧办公等业务场景；研发设计、生产制造领域的数智化水平则较为滞后，仍然处于能力建设阶段。

表 1 各行业数智应用落地的头部场景

行业	研发&设计	生产&服务	营销&销售	运营&管理
第一产业				
农业	- 智能育种 - 农业规划 - 农业装备设计	- 农业装备制造 - 农业生产运营 - 农业技术服务	农产品销售	- 农产品市场分析 - 农资管理 - 农业供应链管理
第二产业				
制造	- 产品设计及优化 - 智能仿真及建模 - 知识产权管理	- 自动化生产线 - 供应链优化 - 设备维护管理	- 客户群体分析 - 销售趋势预测	- 库存管理 - 风险管理 - 成本控制
能源	- 电网/管网设计 - 新能源技术研发	- 智能生产监控 - 自动化生产线 - 智能电力调度	- 需求预测 - 定价策略分析 - 客户体验管理	- 服务质量分析 - 风险管理 - 成本控制
建筑	- 建筑设计辅助 - 建筑信息建模 - 设计方案评审	- 智能施工管理 - 工程技术资料管理	- 市场分析 - 设施选址 - 房地产营销	- 供应链优化 - 设施管理 - 成本控制
第三产业				
金融	- 信贷风险管理 - 投资分析 - 交易算法设计	- 智能客服 - 个性化服务 - 智能投顾	- 自动化精准营销 - 营销活动监管 - 客户关系管理	- 安全合规监管 - 绩效评价及优化
电信	- 网络设计及优化 - 电信产品创新	- 服务体验优化 - 智能客服 - 网络运维优化	- 自动化精准营销 - 营销策略制定 - 用户行为分析	- 安全合规监管 - 风险管理及反欺诈
零售	- 门店设计 - 门店选址 - 品牌定位分析	- 库存管理及优化 - 门店运营 - 客户体验管理	- 品牌营销 - 广告投放分析 - 销售数据分析	- 供应链优化 - 绩效评价 - 门店管理
医疗	- 药物研发 - 临床试验设计 - 医疗装备设计	- 个性化医疗 - 智能诊断	- 医疗产品定位分析 - 患者需求分析	- 医疗资源优化 - 医疗风险管理 - 患者信息管理
教育	- 课程开发 - 智能教育工具开发	- 个性化教学 - 智能教辅 - 虚拟教师	- 招生流程管理 - 目标受众分析 - 课程营销	- 学生及教师管理 - 教学质量监控
交通	- 道路规划及设计 - 智能交通系统研发	- 道路监控 - 交通设施运维 - 交通疏导及调度	- 出行行为分析 - 出行流量预测	- 交通运营管理 - 道路风险预警
文娱	- 内容创作辅助 - 内容素材生成	- 个性化内容推荐 - 用户体验分析	- 营销海报生成 - 目标受众分析 - 广告效果评估	- 内容绩效分析 - 舆情分析 - 版权管理
政务	- 城市规划设计 - 政策制定辅助	- 政务服务 - 公共安全监管 - 公共服务管理	- 政策宣发 - 服务推广	- 绩效管理 - 资源配置及优化

行业范畴来看，数智应用已经渗透到各个行业和领域，大部分行业正处在能力建设阶段。依托对近三届行业数智应用星河案例申报企业的统计分析，整理出各行业数智应用的头部落地场景。各行业数智应用的头部场景存在差异，其中营销及经营管理部分关注的细分场景相似度较高，研发设计、生产服务等场景具备较强的行业特异性，数智应用的侧重点存在较大的差异。总体上来看，以生产型服务业为主导的第三产业在数智应用的深度和广度方面都显著强于其它行业，在数智应用创新过程中发挥着引领作用。

（二） 数智应用落地面临三方面挑战

1. 协同困难阻碍数智应用落地

在大中型企业，数智应用的落地涉及到业务、技术、管理等多个部门的共同协作，跨部门协同难度较大，主要存在两方面问题：**一是缺乏复合型的数智化技术人才**，数智技术的创新应用需要同时精通算法、擅长工程实现、深刻理解业务的复合型人才，目前这类综合性人才相对稀缺，各部门人员跨领域沟通易出现困难。**二是缺乏一体化的顶层设计**，缺乏顶层设计和跨业务、跨领域统筹规划的现象普遍存在，使得对于数智化应用的建设各自为政，难以形成合力，技术和业务需求之间存在脱节，导致数智化应用无法有效解决实际业务问题，造成资源的重复投入和低效建设。

2. 数智应用赋能效益难计量

一方面，对数智应用的产出预估困难，存在着高估短期收益，低

估长期收益的倾向。企业倾向于对大模型等新兴技术短期内能带来的效益抱有过高的预期，期望能够通过快速建设系统、引入工具平台来实现显著的短期收益，但数据质量的持续优化、员工能力的培训、制度流程的适配和企业文化的变革通常需要更长的时间来逐步完善，应用的收益往往需要更长的周期才能显现。**另一方面，数智应用成效评价方式不合理，存在着注重技术指标，轻视业务指标的倾向。**场景类应用的成效评估是一个复杂问题，需要综合考虑多个因素，通常可以细分为技术指标（如算法误差、数据处理性能等）和业务指标（如业务效益变化情况、客户转化率等），由于技术指标较易获取，导致注重技术指标而轻视业务指标的现象普遍，存在着技术和业务需求脱节的风险。

3. 数智应用落地经验难复用

这里将按照数据原生行业（如金融、电信等）和非数据原生行业（如制造业、医疗、交通等）进行探讨。**数据原生行业数智应用所面临的问题主要集中在安全合规方面。**相对其它行业来说，数据原生行业的数据量大，数据质量高。由于涉及到海量个人用户的敏感数据，数智应用过程中面临着数据隐私和安全等问题。**非数据原生行业数智应用所面临的问题主要集中在数据质量、采集成本、多模态处理技术等方面。**非数据原生行业数据量小、质量不稳定且种类复杂，采集成本高、通用性弱，数智应用需解决数据采集成效和处理分析技术适配问题。整体而言，这些行业的数智应用模式尚不成熟，仍需进一步探

索和实践。

（三） AI 驱动数智应用呈现四大趋势

大模型的落地将对企业数智应用模式产生深远影响。具体来看，有以下四大趋势：

1. 大模型技术与小模型协同提升数据分析能力

在大模型技术兴起之前，参数规模较小的模型和机器学习算法在图像分析、社交媒体分析、用户行为分析等分析业务中已取得优异成果。大模型驱动的数据分析工具将趋向于整合多种类型的数据分析能力，如文本、图像、社交媒体、指标及标签等，通过数据智能体（Data Agent）调用高度专业的小模型及机器学习算法并整合分析结果。大模型的应用将有助于企业实现对多模态数据的高质量分析，从而更全面地理解和处理复杂问题。为助力各方开展 ChatBI 等大模型数据分析工具能力建设，中国信通院推出了《大模型驱动的智能数据分析工具技术要求》《数据智能体能力要求》等多项标准，并持续推进相关评测，其中，大模型数据分析工具的交互能力和分析的准确度成为选型的重点方向。

零售领域案例：由于多渠道、多门店、多品牌带来的生意复杂度，在高频营销活动、高频商品上新、高频组织调整带来的冲击下，企业及时获取数据进行决策的诉求变得越来越迫切。观远数据通过大模型赋能的新一代智能分析与决策平台赋能某国内领先美妆公司，协助 10 个部门快速构建起智能数据助手服务，平均不到两周即可引入一个新业务主题，帮助所在部门的业务人员从天级别的数据需求响应周期，优化至分钟级别即可获取所需数据与洞察，需求处理效率提升 200 倍以上。

2. 大模型与知识库融合提升决策能力

作为一种全新的知识载体，大模型和传统的知识库、知识图谱存在较强的互补性：一方面，大模型具备海量通用知识，具备较强的多模态处理能力；另一方面，传统的知识载体则在专业性、可解释性、可靠性方面具备显著的优势，两者结合可实现复杂的知识推理和智能问答，在工业制造、交通物流、国防军事等复杂场景中具备广阔的应用空间。基于专业知识库、知识图谱的检索增强生成（RAG）也成为大模型在垂类场景落地的重要技术路线，在智能问答系统、智能搜索工具、智能客服等场景中具备较大潜力。各领域头部企业开始重视建设完善知识中台、知识工程相关能力，为企业大模型建设提供知识供给的同时也通过大模型构筑面向员工的知识服务，全方位提升企业的决策能力。

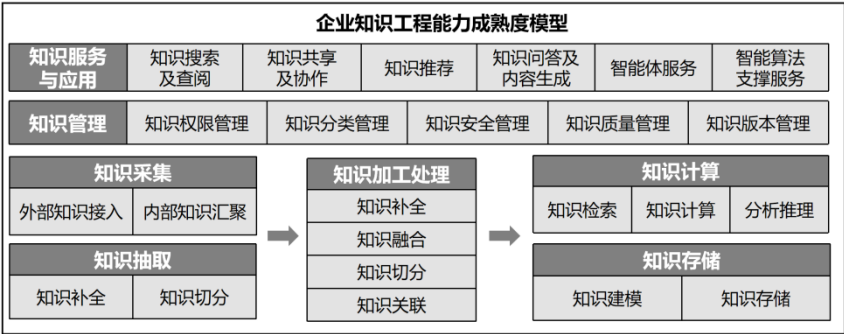


图 10 《企业知识工程能力成熟度模型》标准框架

工业制造领域案例：在某工业制造型企业，大模型通过充分融合企业已有的运行规程、检修方案、工艺流程等知识，形成了该企业独有的行业大模型。该模型以数字专家形式存在，并通过自然语言对话方式帮助专业岗位人员解答专业问题、指导运维操作、审查工作方案、生成经营报告、辅助经营决策，从而提高专业岗位的工作效率与业务创新能力，最终实现提升知识利用效率两倍以上，提升专业人员的创新效率 30%以上，提升企业经营管理决策效率 30%以上。

3. 智能体将逐渐成为大模型应用落地的主要形式

通过建设数字员工能力，并融入企业研发、生产、营销等多环节，企业能够减少人力投入，并实现业务的增长。在研发环节，智能研发助手能精准分析市场趋势，生成创意和设计方案，缩短研发周期，提升创新效率。生产制造方面，运维数字员工可实现生产流程的自动化监控与优化，精准预测设备故障，提高生产效率和产品质量。在营销运营中，运营助手及智能客服能通过数据分析，把握消费者行为，制定个性化营销策略，提升客户满意度和品牌忠诚度。如何构建数字员工平台，提升数字员工的运营能力，探索数字员工和人类员工协作管理模式成为各企业数智化能力建设重点方向。



图 12 《大模型驱动的数字员工平台技术要求》标准框架

能源领域案例：南方电网深圳供电局基于南方电网“大瓦特”体系，研发主网风险巡视“数字员工”。该“数字员工”集成大语言模型、图像识别与知识图谱技术，可自动解析语音指令生成巡视方案、联动视频主站与无人机巢执行特巡任务，实现任务计划秒级自动生成与缺陷智能定级毫秒级精准判定，通过自动化流程贯通风险特巡全环节，系统支持 24 小时实时工单闭环管理、设备健康档案调阅、巡视报告自动生成与工单流转，业务操作环节减少多系统切换，人机协同辅助率达 59%，有效提升主网风险防控效率。

五、 高安全防护护航数据与智能融合新范式

（一） 智能化数据安全治理体系加速升级

政策层面，2024 年来，全球数据安全治理政策法规加速向“精细化”演进。国家数据局成立后，推动数据要素市场化配套政策加速落地，如《数据要素×三年行动计划》进一步深化，促进数据流通与安全协同发展，明确数据分类分级、跨境流动等细则。《网络数据安全条例》于 2025 年 1 月 1 日正式施行，构建起以《网络安全法》《数据安全法》《个人信息保护法》为核心的“三驾马车”监管体系，并细化重要数据目录申报、跨境流动安全评估等要求。国际层面，全球范围内人工智能治理框架逐步成型。欧盟《人工智能法案》于 2024 年全面生效，将 AI 系统风险等级从“不可接受”到“低风险”分为四类，要求高风险系统（如生物识别、关键基础设施）实施严格监管，要求数据训练透明化。美国通过行政令强化 AI 数据安全审查，推动跨境数据流动规则标准化。这些政策既为数据安全治理提供了制度保障，也鞭策企业加快合规能力建设。

技术层面，智能化技术正从“单点突破”转向“体系化渗透”，智能化数据安全技术覆盖率显著提升。在单点能力方面，基于多模态 AI（文本、图像、音视频）的敏感信息检测正在逐步提升准确率，智能化分类分级工具已实现半自动化标签生成，并探索解决非结构化数据分类分级能力。在综合类能力方面，智能化的数据安全运营管控平台成为企业新关注点，在政务、金融等行业已有技术应用实践，支持实

时风险分析、监测与响应。隐私计算在金融联合风控、医疗科研等场景试点应用，保障数据“可用不可见”。

（二） 数据智能浪潮中面临三大安全挑战

1. 安全技术瓶颈仍未突破

面临席卷而来的新场景、新应用引入的安全风险，安全技术虽持续加强，但在应对复杂多变的安全威胁时，仍难以完美应对。从技术层面来看，自动化防护系统应对场景有限、AI 对抗攻防能力失衡、新兴技术融合应用的安全适配性不足等问题凸显，成为数据安全防护体系升级的阻碍。**一是自动化防护难以应对复杂攻击场景。**当前主流自动化防御工具依赖预设规则库与静态策略，面对零日漏洞、新型勒索软件变种等未知威胁时，因缺乏动态行为建模能力，误报与漏报时有发生。尤其在云原生、IoT 边缘计算等异构环境中，传统规则引擎难以覆盖碎片化攻击面，导致防护效能断崖式下降。**二是 AI 对抗攻防能力失衡，防御体系存在技术代差。**在 AI 驱动的攻防博弈中，防御方技术迭代速度滞后于攻击创新。攻击方利用生成式 AI 批量制造深度伪造钓鱼攻击、自适应恶意代码，使攻击效率呈指数级提升，可系统性突破传统安全边界。而防御方的 AI 安全应用仍停留在威胁日志分析、异常模式识别等被动响应阶段。攻防双方在技术迭代速度和资源投入比等方面的不对称性，使得防御方在智能化对抗中长期处于被动局面。

2. 新业态治理复杂度飙升

伴随数据驱动型新业态的爆发式增长，安全管理机制与业务创新速度脱节。管理颗粒度及动态性不足、跨部门协同低效等系统性缺陷导致传统治理框架在敏捷业务场景中效果不佳。一是数智新业态迅速发展但动态管理能力滞后。当前企业安全管理体系难以适应多模态数据的复杂流转特性。文本、图像、音频等多类型数据在业务场景中交叉流动，其敏感度、使用场景差异显著，传统基于静态角色或固定规则的权限管理模式已无法满足动态管控需求。另外，权限策略更新速度远低于数据接口增长量，“僵尸权限”不能及时回收，形成持续性泄露隐患。管理的组织架构上，跨部门协作机制缺失进一步加剧问题——在数智发展的新领域，IT 部门与业务部门权责边界模糊，权限管理、审计、撤销流程冗长，形成安全管理盲区。二是安全策略与创新迭代速度脱节。智能化驱动型业务的快速迭代与安全策略的后置流程形成结构性矛盾，企业迅速开展数智实践并推出新业务功能（如实时数据分析、AI 客服等），但安全策略制定需经历跨部门评审、合规评估、技术适配等多环节，更新周期长，导致创新业务采用“安全后置”上线模式，安全治理缺乏原生适配能力。这种滞后性导致新业务上线时，访问控制、数据加密等安全策略未完全覆盖新应用数据交互模式的风险点。若数据处理活动涉及到上下游流动，供应链安全管理可能进一步暴露体系缺陷，上下游安全策略将左右数据合作过程能力短板，企业自身缺乏对供应链数据事前、事中、事后的监控管理能力。

3. 安全治理落地实践不足

尽管数据安全领域在政策法规与技术探索上持续推进，但在实际落地过程中，企业缺乏成熟、普适的治理范式，标准规范难以转化为可持续执行的落地实践。从行业视角出发，新兴应用场景的合规标准未形成体系、行业间协同治理模式缺位，导致安全治理存在落地盲区。**企业智能化场景安全治理能力未形成明晰体系。**以生成式 AI 为例，从训练数据集的采集、使用安全，到算法的透明度与可解释性要求，再到内容生成和交互合规性，各个环节都有规范指引，但缺乏明确且统一体系化要求。企业在实际应用中只能“摸着石头过河”，逐步搭建体系，这导致企业面临较高的安全合规风险。

（三） 智能驱动的下一代数据安全呈现出两大趋势

1. AI 安全治理从“合规达标”向“主动治理”转型

AI 安全治理加速从“合规达标”向“主动治理”跃迁。针对 AIGC 等新兴场景的应用逐步落地，人工智能安全治理体系应逐步完善。通过搭建包含数据安全、模型算法安全、应用安全、服务安全及通用安全的全局治理框架，全面支撑大模型全生命周期安全管理，确保其技术应用符合安全与合规要求，实现安全能力与业务创新的同步进化。企业治理模式亦发生结构性变革，安全责任从模型责任部门向全员渗透，通过流程规范性提升，构建覆盖“决策层-管理层-实施层-监督层”的立体化治理网络，最终实现安全能力与业务价值的共生共长。

2. AI 赋能基础技术融合发展

智能化技术正推动数据安全基础能力的精细化升级。在数据分类

分级领域，基于自然语言处理和计算机视觉的多模态识别模型，可实现非结构化数据（如文本、图像、音视频）的自动化标签标注与敏感信息定位，降低人工标注成本并提升分类准确性。动态脱敏、访问控制等核心环节同步进化，通过持续学习用户行为，可生成细粒度权限策略，在保障隐私的同时最大化数据可用性。单点技术能力的智能化跃迁，为数智安全治理奠定基础。

在体系化安全能力方面，人工智能进一步向安全运营全流程赋能，推动“被动响应”向“智能防御”转型。通过整合威胁情报、日志审计、行为分析等多源数据，AI 驱动的安全运营平台将逐步构建企业数据资产的全局风险画像，实现攻击路径预测、异常行为溯源等能力。

六、 总结和展望

近年来，数据智能产业蓬勃发展，已成为推动数字经济创新与增长的核心引擎。在数据领域，各行业企业积极投身高质量数据集建设工作，为上层模型训练和应用开发筑牢根基；在算法领域，以大模型为代表的人工智能技术加速渗透，诞生出 GPT-4o、DeepSeek-R1 等标志性成果；在应用领域，智能体的出现进一步降低了数智技术的应用门槛，推动数据智能应用在更广泛的行业及场景中落地生根；在安全领域，在政策规章要求与技术应用突破的双重驱动下，智能化数据安全治理体系加速升级，为数据与智能融合新范式保驾护航。

展望未来，数据智能产业必将持续焕发蓬勃生机，保持强劲的发展势头。随着技术的不断突破和创新，数据智能将进一步提升数据处理与应用的效率 and 安全性，同时，加速向医疗、农业、能源、交通等传统行业渗透，深度赋能各领域数字化转型。数据智能产业将进一步解放个体生产力，驱动企业业务创新，重塑生产关系和产业格局，推动全社会实现跨越式进步，助力我国抢占未来发展制高点，开启智能化发展的崭新篇章。



大数据技术标准推进委员会

地址： 北京市海淀区花园北路 52 号

邮编： 100191

邮箱： TC601@CCSA. org. cn

网址： [www. tc601. com](http://www.tc601.com)