

数据要素发展报告

(2025 年)

中国信息通信研究院云计算与大数据研究所

2025年10月

版权声明

本报告版权属于中国信息通信研究院，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”。违反上述声明者，本
院将追究其相关法律责任。

前 言

党的二十届三中全会强调，要健全因地制宜发展新质生产力体制机制，健全促进实体经济和数字经济深度融合制度。2025 年政府工作报告进一步提出，要激发数字经济创新活力，持续推进“人工智能+”行动，加快完善数据基础制度，深化数据资源开发利用。随着以人工智能为代表的新一轮科技革命和产业变革深入发展，数据作为关键生产要素的战略价值愈加凸显。数据与智能的深度融合，不仅是催生新产业、新模式、新动能的重要力量，更是我国在全球数字经济竞争中构筑新优势、把握发展主动权的关键所在。在此背景下，统筹发展和安全，全面深化数据要素市场化配置改革，持续扩大高质量数据供给，推动数据在经济社会各领域的深度应用，对建设数字中国、推进中国式现代化具有深远意义。

当前，推动数据“供得出、流得动、用得好、保安全”串联起了我国数据要素市场化配置改革的长期目标，在制度、技术与产业的协同向前中，数据要素发展已进入体系化构建与规模化应用的新阶段。一是数据为人工智能发展筑基，推动智能涌现，同时人工智能也使得数据要素价值发挥得到升级拓展，为数字经济发展提供重要引擎。国产大模型的爆发式演进，使高质量、多模态、专业化数据的研发需求空前迫切。数据采集、数据标注等环节与人工智能产业结合为紧密的生态体系。同时，大模型使数据成为能够承载多样化内容的语义载体。智能化数据治理有效降低了海量非结构化数据的处理成本，提升了数据质量。“数据驱动智能，智能优化数据”为数字经济发展筑牢根基。

二是数据制度体系建设全面提速，为数据产业规范发展和数据市场培育筑牢基础。公共数据资源开发利用“1+3”政策体系的相继出台实施，成为破解数据供给瓶颈、激发市场活力的关键突破口。培育全国一体化数据市场的系列政策正加速制定，例如，《国家数据基础设施建设指引》为搭建支撑全国数据资源互联互通的技术底座提供顶层设计，推动各技术路线从理论走向实践；《关于促进数据产业高质量发展的指导意见》则聚焦数据产业发展现状和市场需求，优化布局、加强激励，推动培育多元经营主体。

三是产业实践探索多点开花，在多元创新场景中加速落地。各类主体在确保安全可控的前提下，着力打通数据供给和流通堵点，参与数据要素价值创造。在人工智能的助力下，数据应用向深层次、多领域拓展，持续赋能经济社会发展。数据资产的经济价值在各类应用中加速显化，对于新方法、新路径的探索持续活跃。数据安全治理也从基础的合规达标转向与业务深度融合的主动治理，覆盖数据全生命周期的智能防控体系正在形成。

在此背景下，《数据要素发展报告（2025 年）》进一步讨论了数据要素的理论认识，系统梳理了过去一年我国数据要素发展在供给、流通、应用和安全等方面的最新动态，并在总结发展特征、研判核心挑战的基础上，对未来发展趋势提出展望，以期为各方深入推进数据要素市场化配置改革、全面释放数据要素价值提供有益参考。

目 录

一、智能时代数据要素再认识	1
（一）数据是人工智能发展的核心要素之一	1
（二）人工智能助推数据的作用方式升级拓展	4
（三）数据要素价值释放以全过程贯通为根基	6
二、“供得出”：高质量数据供给水平持续提升	8
（一）制度进展：一系列政策举措推动强化数据供给	9
（二）技术热点：多模态数据研发治理能力寻求突破	13
（三）产业热点：重点领域加快建设高质量数据集	16
三、“流得动”：全国一体化数据市场培育平稳推进	20
（一）制度进展：全国一体化成为顶层设计关键目标	20
（二）技术热点：数据流通利用基础设施建设启动	22
（三）产业热点：数据交易向规范化多样化发展	24
四、“用得好”：多层次应用拓宽数据赋能路径	28
（一）制度进展：“数据要素×”多举措推动数据使用复用	28
（二）技术热点：数智技术融合推动应用范式变革	30
（三）产业热点：数据资产化探索丰富数据价值领域	32
五、“保安全”：数据智能紧密结合共筑安全堡垒	34
（一）制度进展：数据安全治理向 AI 领域延伸	34
（二）技术热点：数据安全技术愈加精准智能	36
（三）产业热点：安全治理从合规达标转向主动治理	37
六、发展展望	38
（一）容错免责机制设计需成为改革关键对象	39
（二）三大引导方向或将在短期内快速推进	41
（三）夯实企业能力仍是发挥市场主导作用的基础	42

图 目 录

图 1 数据要素的三类微观作用方式 5

图 2 数据要素的三次价值释放 7

图 3 历年地级及以上数据开放平台数量增长情况 11

图 4 高质量数据集建设方法论 19

图 5 企业知识工程能力体系 32

图 6 2024 年数据资源入表趋势 33

一、智能时代数据要素再认识

数据作为形成新质生产力的关键生产要素，以其独特的价值增值方式促进科技革命和产业变革，提升全要素生产率。2025 年，由国产大模型深度求索（DeepSeek）引发的新一轮人工智能发展热潮中，数据作为“智能引擎”的作用进一步凸显。2025 年 8 月，国务院印发《关于深入实施“人工智能+”行动的意见》，系统布局“人工智能+”6 大重点行动和 8 大基础支撑能力，对于数据的供给、开放、应用、安全等也提出了相应要求。与此同时，多模态大模型、具身智能等人工智能技术的爆发式演进，也将对数据的产生、治理、流转和利用方式产生深刻影响。因此，应不断深化和完善对数据要素及其价值作用规律的理论认识，以推动数据要素价值释放，赋能智能经济蓬勃发展。

（一）数据是人工智能发展的核心要素之一

数据是人工智能技术和产业发展的重要生产要素。生产要素是对某一时期经济发展中所需重要资源的科学抽象，是对生产过程中所投入成本的高度凝练，数据要素就是在数字经济语境中对数据的指代和对数据价值的强调¹。数据既是人工智能技术迭代和产品研发的关键输入，也是人工智能产业的生产源头和消费对象。技术层面，随着大模型技术应用的快速发展，人工智能的研发重点正从“重点优化模型架构”转向“模型与数据协同优化”，其中高质量数据的作用日益凸显²。当前，从数值记录到行为轨迹，从图像语音到文本知识，万事万物不断转化成数据并“喂”给模型，充当神经网络提取特征的“养

¹ 中国信息通信研究院，《数据要素白皮书（2022 年）》。

² 中国信息通信研究院等，《高质量数据集建设指引》。

料”。输入数据的质量直接影响人工智能模型的训练效果和决策生成的准确性。提高数据质量成为智能水平进一步提升的“先手棋”。**产业层面**，2024 年，全国数据生产总量达 41.06 泽字节（ZB），同比增长 25%。其中，用于开发、训练和推理的数据量同比增长 40.95%，智能家居、智能网联汽车等智能设备数据增速位居前列，分别为 51.43%、29.28%³。数据采集、数据标注、高质量数据集建设等产业与人工智能产业逐渐结合为关联紧密的生态体系，训练数据流通、合成数据供给等业态正不断补足人工智能产业链条。

大模型突破了 DIKW 模型，重塑数据与知识、智慧间的转化关系。

数据的概念及其扮演的角色随着社会变迁而不断变化，在快速变化的技术环境中，数据的功能定位也不断涌现出新的指征和意义。20 世纪 80 年代，“数据—信息—知识—智慧”

（Data-Information-Knowledge-Wisdom, DIKW）模型提出了一种层次分明的认知路径，其中的“数据”侧重指代无意义的原始记录，可被直接输入到计算流程中。大数据和人工智能正是依靠大规模数据计算产生大量人类理性难以直接感知到的信息，为数据赋予意义和解释。这种抽象和提炼式的线性升级路径同样适用于大多数从信息到知识、从知识到智慧的演变过程，但诸如因果、默会知识等知识形式或直觉感知、善恶是非之心等智慧表现的产生并不严格遵循这一线性升级过程。然而，大模型的出现表明，数据可以打破 DIKW 模型中的线性层级壁垒，数据与知识、智慧之间可以直接形成动态交互结构。具体

³ 数据来源：全国数据资源统计调查工作组，全国数据资源调查报告（2024）。

表现形式如，预训练模型将海量数据的关联关系压缩存储，成为输入给机器的经验知识；向量空间中长距离的数据点被自注意力机制敏锐捕捉，效仿人类智慧发掘出看似无关、实则关联的元素；以专业领域信息内容为基础的模型微调优化，生成的内容既可以升级为知识图谱节点，又可以重构训练数据分布。将大模型当作新型数据库和新型反馈系统来使用，既可以从中提取尚未存在的文本、图像，又可以让机器更加积极主动地处理人对数据、信息、知识的多样化要求，从而高效满足技术产业的新趋势、新需求。

例如，**合成数据**拓展数据供给范围，除利用统计分布、模拟仿真、深度学习、强化学习等方式合成特定属性、极端条件、实时交互等数据外，大模型还可自动补充监督标签、思维链过程等数据，补足由于物理限制、伦理约束、安全保护等造成的特定数据缺口。**智能化数据治理**将进一步普及，大模型不仅直接帮助企业降低非结构化数据的处理成本；还能简化清洗流程、准确剔除噪声，显著改善数据质量与一致性；更能有效促进多模态数据关联融合，提高数据调用和研发效率。**知识工程**将再迎发展机遇，作为连接主义技术路线的有益补充，符号主义催生的知识图谱技术在大模型的辅助下，可以自动从复杂的知识库中精准提取实体关系，从而将相对模糊的、原先难以处理的行业经验转化为可计算的知识网络。借助大模型工程化处理的知识库，不仅可以依靠语义理解让知识检索分类更灵活，更是将“静态仓库”变成“动态接口”，直联前端智能应用。

智能时代，各类基于数据的算法应用也将全面嵌入生产生活各领

域，带来更深远的社会治理变革。一方面，数据技术发展的不平衡、数据应用的营利趋向等已带来数据鸿沟、数据垄断、数据霸权等，政府的引导举措和社会的自发行动正在逐步调和矛盾、探寻最优解。另一方面，自工业革命以来，人类不断利用工具，依靠计算逻辑去追求机械可复制、可控的生产能力，逐渐习惯以速度、效率、强度等标准去衡量价值，也逐渐成为现代工具体系的一部分。随着数据向各领域渗透，整个技术体系从控制生产向控制社会生活各方面演进。推荐算法已介入人的偏好选择，大模型也将介入人的思维过程。由此引发的思维封闭化、碎片化、扁平化已初见端倪，人类越来越难以避免地陷入算法预定好的技术系统和社会运转系统。因此，在顺应智能化发展浪潮的同时，如何协调人与机器的关系，确保掌握社会生产秩序的主动治理权，已成为有待探索的深刻命题。

（二）人工智能助推数据的作用方式升级拓展

数据要素的价值无法孤立存在，而是要以产业需求为导向，在与各类业务的深度融合中彰显。从业务功能看，数据作为生产要素的价值本质主要体现在，利用数据的可计算性来降低研发、生产、经营、调度等各业务场景的不确定性、增强稳健性、增加有效信息密度，从而更好实现降本、增效、优化体验等业务目标。从这一角度出发，人工智能是数据发挥作用的重要场景，并非唯一场景。不过，人工智能的发展切实拓展和升级了数据要素的微观作用方式。

具体而言，数据要素的微观作用方式主要有三类（如图 1）。一是准确匹配作用，本质是对逻辑学“同一律”的追求，通过对数据是

否相等的计算，确保数据被准确传输、查询、核验，防止信息损耗。例如，一体化办公、一致性校验、一站式查询等，对提高工作效率、降低潜在风险均有立竿见影的效果。**二是关联预测作用**，本质是对认识论“经验主义”的运用，通过对各类历史数据、相关数据进行统计、回归、分类、聚类等，描述和诊断形势、把握和洞察规律，进而预判未来的事件或行为，提出完成预定目标的方案，极大扩展可供人或机器决策参考的信息量。关联预测作用是挖掘数据价值的关键途径，在用户画像、智能推荐、科学研究、精细化管理等方面均得到广泛应用。**三是重构生成作用**，本质是系统论“涌现”的体现，通过对大量数据汇聚融合、深层计算、解构重组，形成具有极强泛化能力的信息捕获和生成机制。当前，大模型催生的智能应用正在转化为解放人类脑力劳动的新质生产力，这也使得数据的重构生成作用得到凸显。



来源：中国通信标准化协会大数据技术标准推进委员会（CCSA TC601）

图 1 数据要素的三类微观作用方式

智能时代下，数据要素的重构生成作用将迎来更大发挥空间，但

也对数据资源的汇聚、治理、应用、安全合规等提出更高要求。上述三类微观作用中，准确匹配和关联预测作用的发挥大都可以由传统数据交互、分析和决策式人工智能技术提供支撑。准确匹配作用的发挥往往朴实而有效，如银行风控场景中的黑名单查询核验、政务服务“一网通办”等。合理合法的数据关联预测已经给众多业务带来变革，如金融、互联网等行业已普遍利用大数据分析技术，极大改变了客户细分、内容推荐、广告投放、财务健康状况预测、贷款风险评估等业务的形态。相比之下，随着人工智能研发提升投入的不断加大，特别是生成式人工智能的发展加快，数据的重构生成作用将得到进一步的凸显。但与此同时，人工智能训练对于“数据”原材料的质量要求也将不断提升，数据资源的内容丰富性、完整度、标准化程度和对数据进行加工处理的方式与安全合规性等都是影响数据要素作用发挥的关键因素。

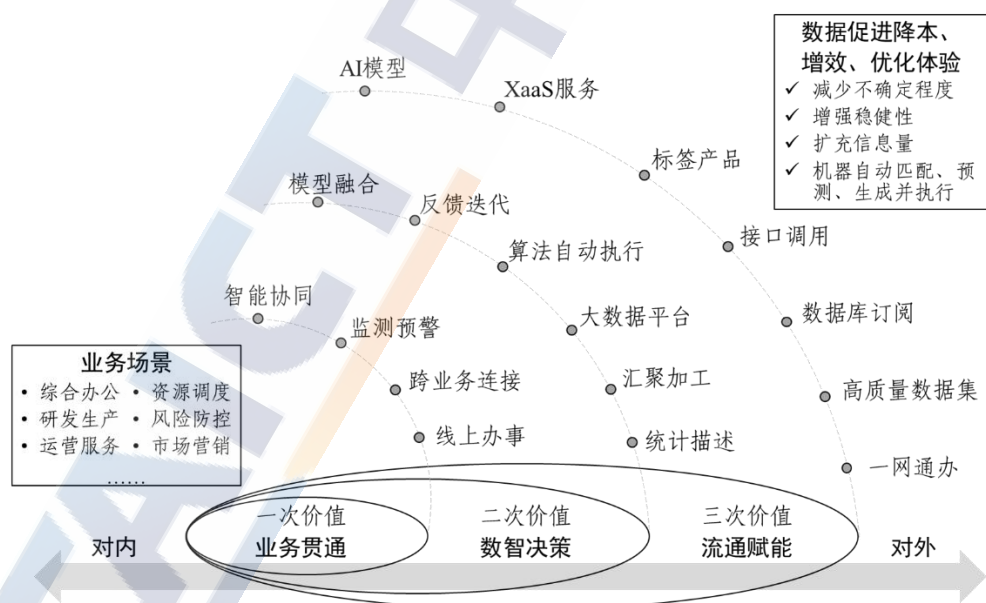
（三）数据要素价值释放以全过程贯通为根基

发展数据要素，推动数据要素市场化配置的根本目的，就是要使数据在各个产业领域不断发挥其应有的作用，从而赋能实体经济，繁荣产业生态。从宏观路径看，数据要素往往经历三次价值释放过程（如图 2），由单一组织内部逐步向更多产业主体扩散。

一次价值体现在数据支撑各组织内部的业务系统运转，实现业务间的贯通。数据随着各业务系统产生，并通过计算机读写实现业务的初步标准化、自动化，打破线下与线上的界限、打破业务流程间的界限。当数据在系统中运转起来时，就已经在生产活动中释放出价值。

二次价值是以存储沉淀地业务数据为基础，以优化主观决策或系统决策为目标的数据复用。生产经营中有大量活动是选择行动方向、决定执行时机、调配资源投入、处理反馈结果，即广义的决策。对组织内部的数据进行二次利用，可以构建出理解、预测乃至控制事物运行的新体系，使决策更加精准智慧。二次价值还可回馈一次价值，深度嵌入到业务系统的正常运转中，保障业务的稳定。

三次价值释放使数据流通到更被需要的地方，让不同来源的高质量数据在新业务场景中融合，实现双赢、多赢的价值再创造。历史经验表明，生产要素的价值实现并不完全依赖流通，但合理的流通有助于充分释放生产要素的全部价值。数据要素具有规模报酬递增效应，越大规模、越多维度的数据所创造的价值越大，对二次价值和一次价值生产单位的综合回馈乃至社会福利也越多。



来源：中国信息通信研究院

图 2 数据要素的三次价值释放

从实现方式来看，数据要素的三次价值之间存在递进关系，后者

往往以前者为基础，对各组织的数字化转型基础、数据治理和应用能力等提出前提要求。但是，从重要性来看，三次价值释放是同等重要的，不能过度强调某一次价值，而是要尊重实际的发展目标和能力基础。产业内各主体发展阶段各不相同，部分主体能够将数据和大模型等前沿技术深度融合，有效实现业务需求，部分主体可将自身积累的优质数据资源加工成产品和服务赋能外部合作方，均能实现各自的产业目标。因此，激活数据要素的战略性价值的根本在于**全面推动各次价值充分释放，将数据以多样、创新的方式投入于经济社会发展全过程**。

进一步地，从数据要素市场化配置改革的主线出发，数据跨主体复用引起供给和需求的对立统一，如何平衡各方对于维护数据安全、维持数据价值、降低摩擦成本等要求，实现数据要素流通赋能的第三次价值释放，对于发展数据要素来说既是目标又是抓手。需要从数据流通的全过程切入，通过一系列制度基础、技术支撑和产业实践的同步推进，最终实现数据“供得出”“流得动”“用得好”“保安全”。

二、“供得出”：高质量数据供给水平持续提升

我国数据资源规模庞大，根据《全国数据资源调查报告（2024）》披露的数据，2024 年，全国数据生产总量达 41.06 泽字节（ZB），同比增长 25%。但是，数据要素的特殊性决定了其价值释放不在于存储，而在于市场化配置牵引下的充分流动和使用。数据供给的规模和质量直接决定了要素价值发挥的基础水平。随着近几年各界对于数据要素重视程度的提升，对数据开发、管理、标注和对外提供等工作保

持积极投入，我国数据供给水平正在逐步提升。

（一）制度进展：一系列政策举措推动强化数据供给

2024 年以来，围绕数据供给相关的政策举措加速出台。一方面，围绕公共数据资源开发利用的“1+3”政策体系初步形成，为扩大公共数据资源供给、引领带动全社会数据资源融合，提供了工作指引；另一方面，面向人工智能发展需求，各地区各部门加快推进高质量数据集建设工作。

公共数据资源开发利用通过三种方式扩大数据有效供给。我国公共数据规模庞大，蕴藏着巨大的政治、经济、社会、文化和生态价值，具有集聚程度高、管控力度大等特征。在数据要素发展初期，需要通过制度设计大力推进公共数据资源的开发利用，这既是促进公共数据供给公平普惠、促进公共数据价值惠及全体人民的重要举措，又是引领激活全社会数据要素潜能、推动数据要素市场化配置改革的关键突破口。2024 年 9 月，中共中央办公厅、国务院办公厅印发《关于加快公共数据资源开发利用的意见》，明确提出通过政务数据共享、公共数据开放、公共数据授权运营三大着力方向。

政务数据共享有法可依。多年来，我国始终致力于持续提高政务数据共享水平。近 5 年各地区政务服务平台数据共享累计超过 5400 亿次，电子证照共享服务累计超 108 亿次⁴。上海汇集公共数据超 2000 亿条，数据接口累计调用超过 480 亿次，并通过政务数据精准画像加快惠企政策“免申即享”等服务⁵。重庆“一表通”改革累计开通用

⁴ 数据来源：国家数据局，《数字中国发展报告（2024 年）》。

⁵ 数据来源：新华社，《政务数据共享迈入法治化新阶段 上海“两张网”释放数字治理新效能》。

户 19 万个，清理历史报表 19 万余个，基层报表数量、处理报表时间和参与报表工作人员均实现 70%及以上缩减⁶。2025 年 6 月，《政务数据共享条例》公布，标志着政务数据共享工作迈入法治化新阶段。该条例首次以行政法规形式明确政务数据共享的总体要求、管理体制、目录管理要求、共享使用要求、平台支撑体系、保障措施和法律责任，构建起“一盘棋”统筹、“一本账”管理、“一站式”服务、“一体化”平台的制度框架。《政务数据共享条例》的实施将进一步厘清权责边界，确立无条件共享、有条件共享、不予共享的分类实施原则，为深化跨层级、跨部门数据融合应用，提升公共服务效率和数字政府建设提供有力支撑。

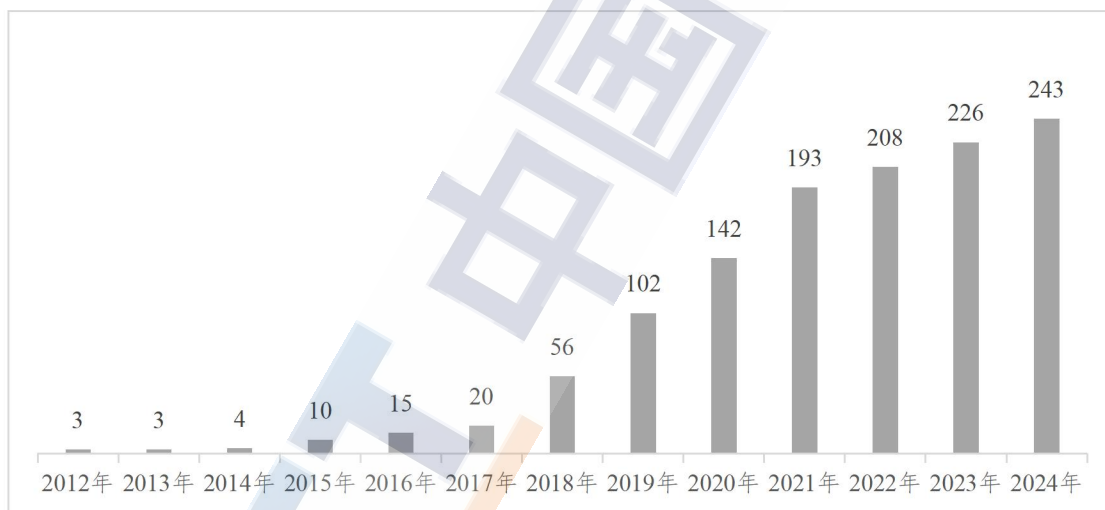
相较于区域，各行业层面的公共数据开放制度建设仍需完善。我国地方政府的公共数据开放已推进多年，通过搭建开放平台和出台地方政策等促进公共数据的公益性供给，很多地方已将公共数据的共享开放纳入绩效考核。截至 2024 年 7 月，我国已有 243 个省级和城市的地方政府上线数据开放平台（如图 3），各地平台上开放的有效数据集达 370320 个，与 2023 年下半年相比，新增 17 个地方平台，平台总数增长约 8%，数据集数量增长约 7%⁷。2025 年，部分地方在已制定的公共数据开放相关管理办法的基础上，继续出台工作细则，如 2025 年 2 月，山东省大数据局印发《山东省公共数据开放工作细则》，无锡市政府印发《无锡市公共数据开放实施细则》，均细化了目录管理、评估更新、获取流程、审核职责等要求，鼓励社会主体开发利用

⁶ 数据来源：新华社，《以法治规范数据共享 护航重庆“一表通”改革》。

⁷ 数据来源：复旦大学数字与移动治理实验室，中国地方公共数据开放利用报告——省域（2024 年度）。

公共数据。

相较于地区层面持续推进，行业层面公共数据开放相关的制度设计还有待完善。虽然多数行业主管部门已通过官网“数据”专栏或统一平台等方面提供数据开放渠道，但访问入口分散、形式差异大、可机读复用的标准不一等情况仍普遍存在。2025 年 2 月，国家发展改革委等部门印发《关于开展物流数据开放互联试点工作的通知》，意在打通政府部门、相关企业及业务系统数据，建立健全物流数据分类及交换应用标准规范，形成物流公共数据共享开放清单。此类制度设计值得向更多行业领域拓展。



来源：《中国地方公共数据开放利用报告——省域（2024 年度）》

图 3 历年地级及以上数据开放平台数量增长情况

三大制度抓手保障公共数据授权运营稳步推进。2025 年，《公共数据资源登记管理暂行办法》《公共数据资源授权运营实施规范（试行）》（简称《实施规范》）《关于建立公共数据资源授权运营价格形成机制的通知》三份配套文件陆续发布，为公共数据授权运营的深入实施明确了方向与规制。公共数据资源登记工作虽然不局限于授权

运营，但对纳入授权运营范围的公共数据资源提出了强制要求，以摸清授权运营覆盖的数据范围及产生的数据产品与服务的情况，确保运营规范化、透明化。同时，国家公共数据资源登记平台的上线也促进了供数单位与用数单位间的供需对接。截至 2025 年 8 月，平台可查询到公共数据资源 1.1 万余项、公共数据产品和服务 700 余项。《实施规范》明确“再开发再利用”是公共数据资源授权运营的基本宗旨，对各相关部门的角色及责任进行划分，并要求通过授权运营实施方案和授权运营协议签订等厘清各地方、各行业授权运营探索中的参与主体、授权范围、运营实践、工作成果等，致力于保障全体公民和市场主体的知情权、监督权和参与权。以此为指引，各地区各部门细化完善授权运营机制，《深圳市南山区公共数据授权运营管理暂行办法》更为详细地明确了各部门的职责分工和授权运营工作流程，浙江省数据局公开采购了公共数据授权运营成效评估服务，促进授权运营工作落地见效。**价格形成机制**基于运营机构特殊角色定位，划清收费范围，规范定价程序，一方面解决了不会定价、不敢定价的问题，保障授权运营各参与方的合法权益；另一方面通过设置最高准许收入和上限收费标准等管控机制，防止其形成垄断超额利润。例如，《福建省公共数据运营服务费收费标准（试行）》将公共数据运营服务费分为技术服务费和存算设施费两项。其中，技术服务费按照应用场景的数据使用量计量，对用于公益事业的公共数据应用场景则不收取技术服务费。2025 年 6 月，首批政府指导定价公共数据授权运营产品在福建交付，合同金额超百万元。

除强化公共数据供给外，解决高质量数据集供给规模不足的问题也成为政策布局的主要着力点。2025 年 2 月，国家数据局组织 27 个部委召开了高质量数据集建设工作启动会，加强统筹协调，深化部门协同。各部委依托自身职责，加快推进垂直行业领域高质量数据集建设。例如，国务院国资委围绕交通物流、绿色低碳、金融服务等 10 个领域推进中央企业高质量数据集建设，并已发布首批 30 项优秀建设成果。各地方政府也积极鼓励和引导企业高质量数据集建设落地，一方面组织遴选评审，选拔优秀案例或试点示范，另一方面设计出台针对性奖补政策，以实质资金补贴的形式吸引优质数据供给。

（二）技术热点：多模态数据研发治理能力寻求突破

数据资源供给的能力建设是以应用为牵引的，人工智能的技术爆发带动了对高质量数据供给的需求。虽然应用侧仍以常规数据集为主，但从技术发展趋势来看，针对于多模态数据的研发治理成为技术突破的重点方向。在各类主体持有的数据中，文本、图像、音视频等无预定义格式或模型的非结构化数据占比较高，其背后蕴含的价值也极为丰富。随着技术进步，当前的数据治理技术已经可以有效支撑各主体可对非结构化数据进行内容识别、提取和解读，并通过提取非结构化数据的元数据与业务、主题标签等信息，将非结构化数据进行结构化。以此为基础，各主体可借鉴以往对结构化数据管理的经验，对多模态数据进行高效研发和治理，为提高数据质量和供给水平提供有力支撑。

多模态数据的研发与治理工作着重关注多种模态数据间的内在关联。随着人工智能技术的迅猛发展，多模态数据的融合水平愈加影

响数据价值释放的最终效果。例如，自动驾驶场景中，需要融合摄像头影像与雷达信号进行综合判断；智能客服场景中，文字、图片描述与通话音频也需要归一化处理。面对这一态势，很多企业秉承全域数据治理理念，对不同模态的数据制定统一的数据生命周期管理规范，既保障数据在采集、处理、存储、应用等环节中的质量与安全，又利用知识图谱等技术将企业所有数据资源进行梳理和关联。在此基础上，企业将革新数据研发治理范式，打通数据供给与应用的桥梁。

数据研发运营一体化（DataOps）将敏捷、精益等理念融入数据开发过程，通过对数据相关人员、工具和流程的重新组织，打破协作壁垒，构建集开发、治理、运营于一体的自动化数据流水线，不断提升数据产品交付效率与质量。企业基于 **DataOps** 理念，建立融合协同的数据组织，打造一体化的数据平台，并不断深化数据管理与应用的文化，逐步构建起敏捷、高效、标准、健康的数据管理与运营体系，实现从数据“小作坊”到“大工厂”的转型。我国已有一批数字化领先企业通过引入和延伸 **DataOps** 方法，开展了对结构化数据、半结构化数据、非结构化数据等多种模态数据的工程实践。这一举措不仅实现对原有面向 **BI** 应用的数据交付效率和质量的提升，还达成了对 **AI** 应用的高效、高质量与安全供数。实现了数据质量、效率和成本的大幅优化。

数据编织（Data Fabric）是一种以“逻辑集中、物理分散”为核心思想的数据管理架构，将全域数据统一管理。它以元数据和事件驱动为基础，具有分布式、自动化、智能化、民主化等特征。数据编织

将多种技术进行有机融合，实现对多源异构数据的动态集成。用户在使用数据管理和应用能力的同时，解决企业处理多源异构数据时高成本和低效率问题。从管理价值看，数据编织通过简化数据集成和数据处理过程，提供更快速的数据访问和分析服务，从而帮助企业更快地做出决策，更快地响应市场需求。从经济价值看，数据编织能减少数据冗余、简化数据集成的复杂性，从而实现存储成本、技术成本和运营成本的降低。从社会价值看，数据编织可以连接和管理市场主体中多源异构的数据，推进数据按需使用和计价，支撑数据要素流通，促进数据再开发再利用。

智能化数据治理旨在运用人工智能技术开展数据治理工作，帮助企业的数据治理工作实现从劳动密集型向技术密集型工作转变。智能化数据治理不仅是对传统治理流程的简单“提速”，而是基于 AI 认知能力对治理范式进行的一次范式升级。它将过往以人力为核心的“流程驱动”模式，转变为以模型为核心的“认知驱动”模式，使企业数据质量从“劳动密集型”跃迁至“技术密集型”，并带来三重价值跃迁。一是从“规则+人工”到“模型+自主”，效率与质量双升。传统治理高度依赖人工编写规则、人工抽检、人工补录，既耗时又难以覆盖全量数据。智能化治理通过语义理解、逻辑推理与智能生成，把规则沉淀为可自我迭代的模型。二是从“割裂烟囱”到“协同飞轮”，组织协同闭环。传统治理工具分散、口径不一，导致“业务-数据-技术”三方信息孤岛。智能化治理平台以统一语义层为底座，把需求、标准、问题、评估结果全部沉淀为可解释、可复用的知识图谱。三是

从“单模态”到“多模态”，释放全域数据价值。传统治理聚焦结构化数据，对图像、语音、文本、时序、图数据等多模态资产缺乏有效手段。智能化治理借助跨模态大模型，打通“感知-认知-治理”链路。由此，企业可将原本沉睡的非结构化数据纳入统一治理框架，实现跨模态数据的可发现、可理解、可信任，为多业务场景（风控、营销、制造、合规）提供全域、高质量的数据燃料。

（三）产业热点：重点领域加快建设高质量数据集

高质量数据集是经过采集、加工等数据处理，可直接用于开发和训练人工智能模型，能有效提升模型性能的数据的集合。高质量数据集不仅是供给人工智能模型训练、推理和验证的基础，更是人工智能赋能行业的核心驱动力。在这个意义上，高质量数据集不仅关注传统数据质量管理中规范性、完整性、准确性、一致性、时效性、可访问性等方面的质量，还进一步关注类别全面性、维度均衡性和内容安全性等指标，以满足人工智能模型训练和应用的需求，推动人工智能产业的健康、可持续发展。由于与人工智能模型训练紧密相关，高质量数据集的分类体系也相对复杂。从数据模态看，可以分为单模态数据和多模态数据。从流程阶段看，可以分为预训练数据集、微调数据集和评测数据集，其核心逻辑是“先通识教育，再专业精修”。从训练方向看，可以分为通识数据集、行业通识数据集和行业专识数据集，其类型划分主要通过数据集的知识内容、来源类型、时效性、标注人员类型、敏感程度、模型类型、主题范围等维度进行。

高质量数据集建设进入加速期，重点行业领域积极探索实践创新。

2024 年 12 月，国家发展改革委等部门联合印发的《关于促进数据产业高质量发展的指导意见》首次明确提出“高质量数据集”，将其作为人工智能与实体经济融合的核心载体，并提出开发行业数据集的具体要求，数据集高质量发展成为行业发展重要目标。目前，面向工业制造、医疗卫生、交通运输等行业领域人工智能训练的高质量数据集供给较为集中，低空经济、具身智能等创新领域也因产业发展驱动陆续涌现。工业制造场景数据集源于采集端存在大量设备采集的图片、音频、视频等非结构化数据，多应用于企业数字化转型的智能化场景，如生产制造、故障诊断、智能运维、设备状态监测等。医疗卫生场景数据集是在传统专病数据集的基础上，结合影像数据、临床科研数据、医学术语、药品说明书等多模态数据进一步加工形成，主要应用于辅助诊疗、辅助决策场景。交通运输场景数据集是通过对车载传感器、路侧设备、卫星定位、交通管理系统等多源数据进行加工标注而形成，以服务智能驾驶、智慧交通场景为主。低空经济场景数据集通过对海量飞行数据、物流数据、地理信息数据等的加工融合，优化低空空域管理、提高飞行安全性、提升物流配送效率，为低空经济快速发展提供有力支撑。

当前，高质量数据集供给方大多集中于开源社区、数据服务企业、数据标注基地和数据交易场所。在开源社区方面，魔搭、飞桨、天池、帕依提提、超神经、智源、和鲸、启智、聚数力等平台提供多类型公开数据集，适用于基础的人工智能模型任务。在专业数据服务商方面，海天瑞声、数据堂等代表企业在原有业务的基础上进一步拓展丰富高

质量数据集产品。例如，海天瑞声在其年报中列示了高质量通用图文数据、通用场景视频无描述数据集等数据资源的账面价值，直观展示了该公司作为我国领先的人工智能训练数据专业供给方的能力。在数据标注方面，四川成都、辽宁沈阳等 7 个标注基地已形成上百个高质量数据集。此外，多地数据交易场所也陆续建立高质量数据集专区，汇聚多模态优质训练数据。例如，北京国际大数据交易所已发布约 300 个高质量数据集，涵盖 10 余个应用领域数据资源地图⁸。

高质量数据集建设的方法论初步成型（如图 4）。高质量数据集建设是一项覆盖数据集全生命周期的系统工程，业界逐步形成两种典型的建设模式：一是“场景驱动”模式。以明确的业务需求或场景为起点，通过“需求拆解—数据设计—数据采集—数据处理—数据质量检测—数据运营”的闭环，确保数据集对场景的智能化水平有所提升。二是“数据驱动”模式。以已积累的大量、多源异构数据为基础，通过主动的数据探索、关联分析与价值挖掘，反向发现潜在的业务需求或优化方向。两种模式相互交替、动态互补，因此在实际建设过程中，**需根据场景和需求执行研发管理、交付管理、运维管理、运营管理四大核心环节**。研发管理是对数据集的生成流程进行管控，覆盖需求管理、设计管理和数据加工三个环节。交付管理是面向数据集的交付过程开展管控活动，主要有测试管理和发布管理两个环节。运维管理是关注数据集的日常监控和维护，对过程中涉及的数据、计算、存储资源进行管理。运营管理关注数据集在用户端的使用情况，衡量投入产

⁸ 数据来源：北京国际大数据交易所累计备案交易金额近百亿[N]. 北京日报，2025-1-21(3)。

出的收益，关注数据集长期的质量提升。在具体的实施过程中，各环节顺序可根据不同业务场景和建设需求调整，也可选择不执行。



来源：中国信息通信研究院

图 4 高质量数据集建设方法论

七项关键技术能力成为高质量数据集建设的核心保障。建设高质量数据集离不开数据采集、数据处理、数据管理、数据标注、数据合成、数据质检、数据服务运营等关键技术能力支撑。这些技术能力可以集成在一个或多个技术工具中。其中，多模态数据处理技术正迎来快速发展的关键时期。一方面，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，将实现更全面、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，多模态数据处理技术将助力车辆更精准地感知路况并做出快速反应。另一方面，多模态模型正朝着高效与轻量化方向发展，多模态模型通过知识蒸馏、参数共享等技术实现性能与效率的平衡，降低计算成本，推动数据处理技术在边缘计算和移动设备上的广泛应用。

整体来看，面向高水平数据供给的一系列工作正在国家统筹下加快推进，在实践中仍存在现实困难与挑战。一是数据治理工作的投入

产出比例失衡，数据清洗、标注等环节所需的成本占比高，但成效不够显性，价值转化周期长，对公共部门和企业的投入持续性提出挑战。二是基础数据的开放汇聚难度大，权属规则不明晰、供给激励机制不明确再加上安全合规边界模糊，使得数据提供方的供给意愿较难提升，无论公共数据授权运营还是跨主体的行业高质量数据集共建，落地事件中均面临较大的沟通协调困难。

三、“流得动”：全国一体化数据市场培育平稳推进

流通赋能是数据要素价值飞跃的关键引擎。我国锚定全国一体化数据市场建设战略目标，通过制度供给与技术设施建设双轮驱动，不断探索数据流通交易过程中安全互信等难题的破解之道，引导广大经营主体稳步拓展市场空间、创新业态模式，构建规范有序且充满活力的发展环境。进入 2025 年，在“全国统一大市场”顶层设计下，数据市场建设的制度框架更加清晰，数据流通利用基础设施的路径初步形成，推动数据流通继续向高效有序、模式创新发展。

（一）制度进展：全国一体化成为顶层设计关键目标

统筹谋划、系统培育成为数据市场相关政策制定的出发点。党的二十届三中全会提出“培育全国一体化技术和数据市场”，从战略高度强调了坚持全国“一盘棋”，促进数据跨层级、跨地域安全有序畅通流动的意义。2025 年 1 月，国家发展改革委印发《全国统一大市场建设指引（试行）》，要求有关部门加快完善统一的数据资源登记规则，制定全国统一的数据交易、安全等标准体系，促进数据要素合规高效流通交易，建立健全数据要素收益分配制度。2025 年 4 月，

国家数据局公布《构建数据基础制度更好发挥数据要素作用 2025 年工作要点》，将完善数据市场规则体系、培育多类型市场主体等作为重要任务，加速全国一体化数据市场培育进程。以规则体系完善为例，国家数据局、市场监管总局制定并发布了四份数据流通交易合同示范文本，以期为数据市场参与主体提供范本和参考，降低参与流通交易的对接成本、提高效率。

多元市场主体培育加速。数据企业是全国一体化数据市场建设的“主力军”。一体化数据市场建设离不开繁荣活跃、竞争有序的市场经营主体。《关于促进数据产业高质量发展的指导意见》《关于促进企业数据资源开发利用的意见》等文件陆续发布，旨在加强产业规划和政策保障，破除企业面临的堵点难点，推动企业主体作用的充分发挥。以此为牵引，各地加快布局。例如，上海、河南等地促进数据产业发展的实施意见或专项政策，江苏、杭州等地设立数据产业发展基金，北京海淀区、杭州高新区（滨江）等地搭建专业服务平台或企业孵化载体，提供数据产业发展所需的生态支撑。

各地方探索推进区域一体化。在国家顶层设计指引下，区域协同发展和综合试验区建设成为培育全国一体化数据市场的制度探索试验田。例如，江苏省数据局开展全省存量数据交易机构整合工作专题督导，要求实现全省共建共用“一个数据交易所、一套数据交易规则体系、一个数据交易平台系统”，整体融入全国一体化数据市场。上海、江苏、浙江、安徽正在探索长三角数据领域一体化工作，建立长三角数据专题合作机制，旨在破除行政藩篱，推进区域一体化。此外，

国家数据局部署建设 10 个国家数据要素综合试验区，在培育经营主体、繁荣壮大数据市场等方面开展先行先试。例如，湖南已印发建设方案，明确以制度建设为主线，以促进数据流通利用为重点，坚持全省“一盘棋”。

相关制度及政策设计的落地效果仍需经时间和实践检验，其自身也需在实践中不断调整完善。当前，数据要素配置机制不清晰、数据市场监管不明确等挑战仍然存在。诸如公共数据、企业数据、个人数据等不同类型的数据的流通边界、数据产权“三权分置”的具体界定标准等核心问题，尚未形成可直接执行的操作规范，大量实践仍需要“摸着石头过河”。因此，在鼓励积极探索的同时，需要注意规避不同地区、不同领域制度规则设计不统一、标准认定不一致的问题，避免全国一体化市场构建停留“框架协同”层面，而是要确保实质贯通。

（二）技术热点：数据流通利用基础设施建设启动

数据流通利用技术发展日趋成熟，为构建数据基础设施提供基础条件。数据的生产要素化过程，离不开通过高流动性发挥其乘数效应。从流动性来看，相较于其他生产要素，数据要素的流动速度更快、流通范围更广、应用程度更深，由此也引发大量数据权属争议、数据安全与合规挑战。经过近年来的探索积累，隐私计算、数据沙箱、使用控制等支撑数据安全流通的单点技术日趋成熟，基本满足了具体场景的业务需求。然而，面向全国一体化数据市场的建设目标，为解决所有数据供需方面临的共性问题，构建适应数据要素特征、有效促进数据流通利用的数据基础设施已迫在眉睫。

数据流通利用基础设施的技术路径初步形成。为贯彻落实党的二十届三中全会关于建设和运营国家数据基础设施，促进数据共享的部署要求，国家发展改革委、国家数据局、工业和信息化部组织制定了《国家数据基础设施建设指引》（以下简称《建设指引》），为我国数据基础设施建设给出了明确的方向和路径。国家数据基础设施在传统网络设施、算力设施基础上，提出构建数据流通利用设施，为跨层级、跨地域、跨系统、跨部门、跨业务数据流通利用提供安全可信环境，包括可信数据空间、数场、数据元件、数联网、区块链网络、隐私保护计算平台等技术设施。其中，可信数据空间、数场、数据元件、数联网均是一种综合性方案或框架，区块链网络、隐私保护计算平台则是单一某种技术进行基础设施化的具体表现。

各关键技术路线既有特定优势，又相互关联，有利于通过技术路线间融合解决实际复杂问题。可信数据空间是数据流通的“安全港湾”，基于数据使用控制、身份管理、数据标识、数字合约等技术服务，构建可信管控、资源交互、价值共创三大能力，实现多方数据的可信连接、共享共用。数场是数据流通的“综合性枢纽”，从点、线、面、场、安全 5 个维度构建标准化技术框架，依托各种技术、各类组件、各项流程机制等密切配合，实现数据安全高效流通。数据元件是数据流通的“中间态”，强调将原始数据与数据应用“解耦”，支持采用标准化工序完成数据产品规模化加工、生产和再利用，适用于大规模数据加工和生产场景。数联网是数据流通的“服务网络”，通过高速率、低时延、弹性扩扩容的数据服务，实现数据就近接入、全网可达，以标准化协

议替代传统专线或平台式连接。区块链网络是数据流通的“信任基石”，支撑其他数据基础设施实现多中心化、共识可信、不可篡改、可追溯等。隐私保护计算平台是数据流通的“安全引擎”，针对高敏感数据跨域流通利用场景，提供开放普惠的隐私保护计算公共服务。

探索初期，先行先试为基础建设最优解积累经验。数据流通利用设施需要一体化打通数据的接入、传输、流通、监测等环节，为大规模高效规范的数据流通利用活动提供保障，为全国一体化数据市场提供支撑。从具体推进路径来看，《建设指引》提出了三阶段的计划，从先行先试开始，逐步收敛和明确核心功能模块，完善顶层设计；再到数网、数算相关设施充分融合，基本覆盖全国大中小型城市；最终实现横向联通、纵向贯通、协调有力的国家数据基础设施基本格局。由于当前正处于起步探索的第一阶段，截至 2025 年 8 月，国家数据局已组织了两批国家数据基础设施建设先行先试工作，力争挖掘和打造一批具有代表性、示范性的数据基础设施建设方案和典型实践。围绕六条技术路线，各试点单位已经提出各自的建设方案，立足本地特色，启动创新探索。

（三）产业热点：数据交易向规范化多样化发展

场外交易是数据流通交易的“主阵地”，但公开披露的信息不够充分。2024 年，全国数据市场交易规模预计超 1600 亿元，其中场内市场数据交易（含备案交易）规模预计超 300 亿元⁹，场外数据交易仍为数据供需方相互匹配的主要途径。数据流通交易并非新生事物，

⁹ 数据来源：国家数据局。

早在 1803 年，一群伦敦裁缝互相交换不能偿清债务的客户信息，成为征信机构益博睿（Experian）发展的基石。在数据交易场所概念兴起之前，数据采购、接口交互、终端订阅等形式的场外数据交易模式就已经形成一定规模，在金融征信、商业信息查询等领域已经形成了相对成熟而固化的市场格局，甚至部分领域存在垄断风险。例如，中国知网深耕学术出版行业数据，深度开展出版机构对接、数据清洗挖掘等工作，构建起庞大的文献平台，并以年度订阅等模式向单位和个人提供数据查询服务。2022 年 12 月，国家市场监督管理总局查明，知网在中国境内中文学术文献网络数据库服务市场具有支配地位，2014 年以来知网滥用该支配地位实施垄断行为。

场外数据交易的公开披露信息仍然不足，导致难以对参与主体、市场价格、成交规模、交易形式等有宏观全面地掌握。从 2024 年上市公司年报对数据资源的披露情况看，有 66 家公司披露了涉及数据产品的情况，一方面反映出某些行业领域已形成较为稳定的数据供需关系，一些企业已深度参与场外数据市场；另一方面也反映出还有大量企业参与数据市场的信息尚未公开。例如，科大讯飞本期增加外购数据资源近 6746 万元，但未详细披露数据供给渠道。中科星图年报显示，其新增遥感共性产品 7 大类 50 个数据集，已形成上中下游全面互通的数据及服务全链路，甚至在 5 个数据交易所上架产品，但其报表中的“数据资源”项为 0，无法得知交易价格、规模等情况。当然，数据供给渠道、数据产品报价等一定程度上涉及企业的商业秘密，并非必须披露的内容。但适当披露也将有助于提升透明度，促进市场

繁荣，因此还需探索相关信息的披露方式和尺度。

数据供需方在场外持续强化合规要求。企业为维护良好声誉、规避经营风险、保障业务长期稳定发展，本身就具备一定的合规动力。数据需求方为避免连带法律责任和商业损失，自觉对合规性提出要求，如建立严格的内控合规流程，采购外部数据时做穿透性审查，要求数据供给方提供数据来源合法性证明、相关权利主体授权协议等。例如，商业银行在 2016 年前后即开始制定外部数据管理办法并持续完善，强化合规管理流程。相应地，数据供给方为提高市场竞争力、降低违规风险，也会主动提升合规能力，完善交易流程。随着全国一体化数据市场相关政策的出台，数据供需方正密切关注数据流通交易合同示范文本、数据流通交易规则等动向，以便及时改造调整，确保数据供给、采购等行为符合规范。

数据交易场所继续探索发展路径。数据交易只是手段，面向场景的应用才是最终目的。与其他大部分商品不同，数据作为商品参与交易时，客观上存在较大异质性，不同行业、不同场景对于数据类型、数据格式、加密手段、传输方式、处理应用方式等要求差异大，点对点沟通与个性化供需对接成为数据流通交易的基本特征。因此，各数据交易场所继续探索合规增值服务、引领细分应用场景、推动公开透明等多个方向，从而寻找不可替代性。当前，多个数据交易场所的数据交易规则体系已初步成熟。例如，深圳数据交易所围绕数据交易前、中、后全周期，出台管理办法 19 项、规则指引 32 项、技术规范 9 项，主导或参与标准编制 25 项，推动数据交易更安全、更可信。面向应

用场景打造数据专区逐渐成为主流。截至 2025 年 1 月，共有 26 个数据交易场所建设了数据专区，其中 25 个按推荐场景进行专区建设¹⁰。例如，贵阳大数据交易所已打造 21 个数据专区，覆盖农业、医疗、产业互联网等重点应用场景。此外，在 2025 年数字中国建设峰会上，中国价格协会数据价格专业委员会、贵阳大数据交易所、北京国际大数据交易所、上海数据交易所、深圳数据交易所、福建大数据交易所、安证合规集团共同启动“数价链”全国一体化数据价格信息服务平台，推动数据交易价格互认互通。

打击数据黑市交易仍然任重道远。尽管数据交易合规体系逐步完善，但以非法获取、贩卖公民个人信息为代表的黑市交易屡禁不止，仍然威胁着个人隐私和社会安全。黑市数据交易者突破法律底线，在牟取不当利益的同时，也助长了电信诈骗、网络暴力等下游犯罪，与合法市场中企业努力寻求数据合规路径形成了鲜明对比。不法分子利用非法爬虫、撞库攻击、木马植入等技术手段强行窃取敏感个人信息，众多个人标签被明码标价。更令人忧虑的是，近年来已有多起涉及其企事业单位内部人员等利用职务便利非法查询、窃取并出售公民个人信息的案件被查处，揭示出数据安全管理的内部风险。尽管黑市交易具有匿名性，资金链条难以追溯，导致打击行动面临巨大困难，但数据黑市是释放数据要素价值过程中必须清除的重大障碍，需要不断提高打击水平、加大打击力度，维护市场秩序。

¹⁰ 数据来源：天翼智库，《2024 年我国主要数据交易场所发展回顾（下半年期）》。

四、“用得好”：多层次应用拓宽数据赋能路径

完成数据要素价值释放回路的关键在于打通应用落地的“最后一公里”，在具体场景中发挥出数据要素的乘数效应。只有持续丰富产业实践中的数据应用，才能取得数据价值转化的实际成效。2025 年，“数据要素×”行动继续自上而下推动更广泛领域的数据创新应用，人工智能在技术层面驱动数据应用技术变革，产业实践也在将数据投入具体业务场景、解决实际问题的基础上，积极探索将数据进行资产化处理对于企业经营的赋能路径，拓宽数据要素价值发挥的空间。

（一）制度进展：“数据要素×”多举措推动数据使用复用

“数据要素×”行动激励数据应用场景深化和拓展，展现出数据要素巨大潜力。《“数据要素×”三年行动计划（2024—2026 年）》印发以来，各地区各行业围绕工业制造、现代农业、商贸流通、交通运输、金融服务、科技创新、文化旅游、医疗健康、应急管理、气象服务、城市治理、绿色低碳等十二个场景开展行动，促进数据要素价值落到实处。作为一项自上向下推动数据与实体经济融合应用的行动计划，围绕推动“数据要素×”行动落地，国家数据局和地方数据主管部门在试点建设、案例遴选、赛事活动等多方面开展工作部署。

多部门协同开展试点建设。为推动“数据要素×”走深向实，金融、气象、文物、医药等多个行业主管部门积极酝酿创新试点。在金融领域，中国人民银行将研究制定新阶段金融科技发展规划，出台深化运用金融科技推动金融数字化智能化转型的政策文件，发挥金融海

量数据和多元应用场景优势，增强数据基础能力和底座支撑。中国证监会开展“数据要素×资本市场”专项试点工作，第一批 11 个试点项目牵头单位涵盖了资本市场核心机构、经营机构和信息技术服务商，在证券市场、期货市场、区域股权市场三个领域开展了有益探索。在医药领域，国家中医药管理局布局 12 个数字中医药试点，遴选 127 个单位开展智慧中医医院试点项目建设，重点加强中医电子病历、中药药学服务、名老中医传承、中医临床科研、中医智慧医共体等智慧应用。

案例和赛事为产业实践提供大量高价值参考经验。通过案例遴选和大赛平台识别和推广典型经验是“数据要素×”相关制度和政策落地的重要抓手。截至 2025 年 7 月，国家数据局发布两批共 48 个“数据要素×”典型案例，释放引导效应。2024 年 10 月，首届“数据要素×”大赛落幕，全国共有近 2 万支队伍踊跃参赛，涌现出一大批数据开发利用的新场景、新模式，最终角逐出 12 个赛道共 71 支获奖队伍。2025 年 4 月，2025 年“数据要素×”大赛正式启动，对比 2024 年与 2025 年的变化可以看到，“数据要素×”大赛的办赛重心已从初期的理念倡导和场景探索逐渐转向更侧重市场化价值导向、更强调具体行业问题解决和更注重体系化联动的深化阶段。此外，从各地印发的分赛通知看，各地在设置特色赛道、明确赛题方向、配套激励政策等方面做了因地制宜的精心设计，力求在赛事规模、质量、实效上取得新的突破。

“数据要素×”的务实创新导向有利于避免计划式落地应用。过

往信息化、数字化建设浪潮中，虽然涌现了不少应用场景与应用系统，但由于不少项目采用计划式选择场景落地的方式，缺乏挖掘数据复用潜力的意识和能力，导致大量数据被封闭在各自系统之中，形成了相互割裂的数据壁垒，难以得到统一有效治理，甚至催收大量无人问津的“僵尸应用”。为避免此类问题重演，“数据要素×”行动继续升级组织方式，构建起更广泛的跨部门、跨层级纵横联动机制，新增开放性创新赛道，着重激励全社会参与，在解决行业实际痛点过程中打造“供数—用数—收益”价值闭环，积累数据应用可持续发展经验。例如，2025 年的赛题设置中，增设了更具灵活性的开放性创新赛道，从医疗健康领域细分出了独立的“医疗保障”赛道，每个赛道的赛题也更加丰富具体，从而让数据应用能更精细地应对行业需求，更有针对性地形成可复制推广的方案。

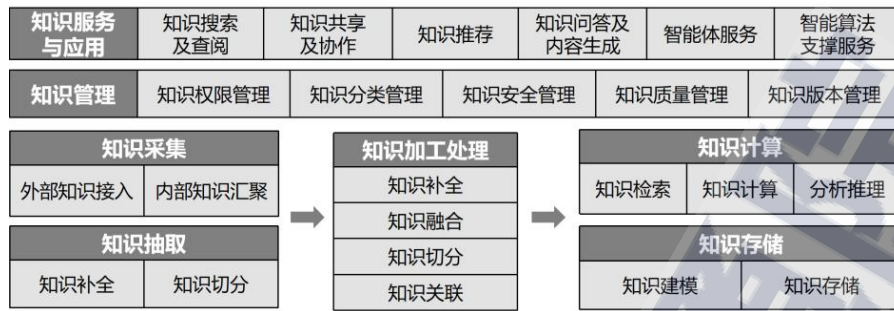
（二）技术热点：数智技术融合推动应用范式变革

数据智能体与大小模型协同，提升数据分析洞察能力。在大模型技术兴起之前，参数规模较小的数据挖掘模型和机器学习算法在图像语义理解、社交媒体分析、用户行为挖掘等业务中的应用已取得优异成果。下一阶段，大模型驱动的数据分析工具将趋向于整合文本、图像、社交媒体、指标及标签等多种类型的数据分析能力，通过数据智能体（Data Agent）自动调用专业的小模型及机器学习算法并整合分析结果，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 and 处理复杂问题。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当前选型的重点方向。

大语言模型、多模态大模型等驱动信息检索类数据应用向智能化飞速提升。信息化时代早期，搜索引擎是公众获得信息的重要渠道。传统搜索依赖关键词匹配，难以解析自然语言中的隐含意图，面对模糊提问或复杂场景时，搜索结果难以满足业务应用需求。大模型赋能的智能搜索技术通过融合自然语言处理、多模态理解与实时检索等方式关联起各种类型的数据，实现了对海量信息的精准筛选与深度整合。例如，众多大模型厂商推出的人工智能搜索工具、智能搜索平台等支持多模态结果输出，显著提升了信息获取的内容质量与用户体验。以大语言模型、多模态大模型为代表的人工智能技术正在同传统信息检索技术快速融合、协同发展，推动信息检索领域的数应用范式从“链接索引”向“智能搜索”转变。

大模型与知识库、知识图谱融合赋能精准决策。大模型作为一种可以通过交互提取文本、图像等的新型数据库，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前景。基于领域专业知识库、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、智能客服等场景中具备较大潜力。目前，各领域头部企业正在着力构建并提升知识中台、知识工程相关能力，从海量数据中抽取企业专有知识体系，既为企业大模型建设提供知识支撑，又借助大模型打造面

向员工的智慧知识服务，全方位提升企业运营决策能力（如图 5）。



来源：中国信息通信研究院

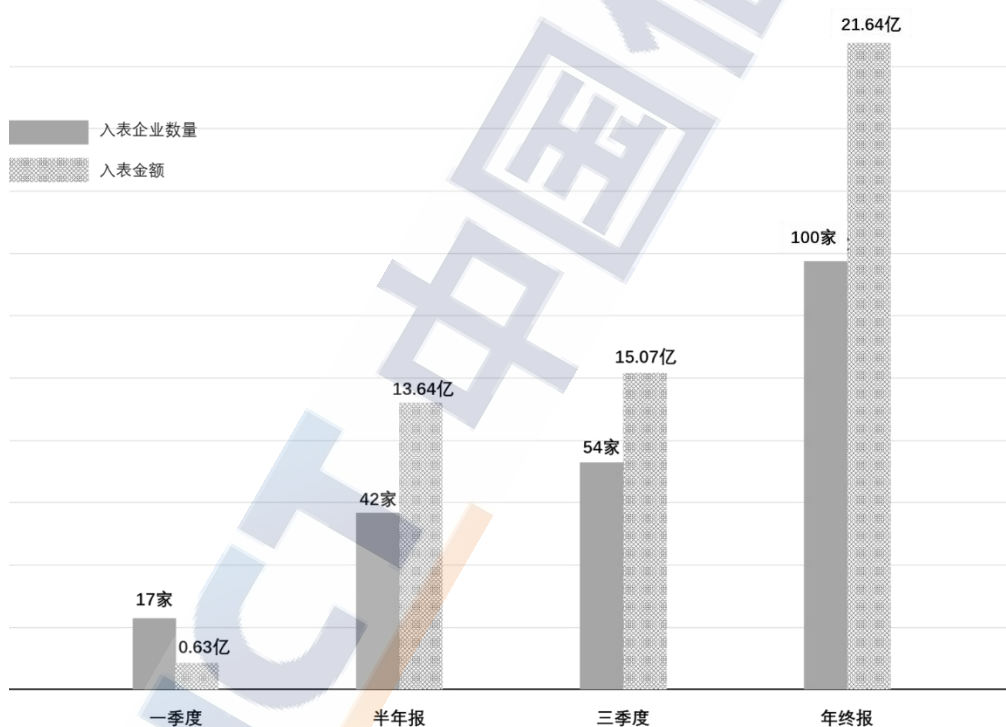
图 5 企业知识工程能力体系

（三）产业热点：数据资产化探索丰富数据价值领域

在具体业务场景中通过数据应用实现提质增效是将数据视为资源的直接应用，而将数据纳入财务报表处理范围或引入金融工具将数据价值“变现”则是将数据视为资产的另一层次应用。将数据进行“资产化”处理和应用，一方面是企业发现和确认数据对业务的贡献程度、促进数据的市场价值“显性化”的一种可操作抓手，另一方面也可以通过财税、金融等手段在一定程度上对企业经营管理进行赋能。

企业数据资源入表继续推进，范围从头部企业向各行业扩展。以上市公司为例，根据各公司年报，2024 年共有 100 家上市公司披露了数据资源入表的相关事项，占全部上市公司数量的 1.85%，数据资源入表总金额为 21.64 亿元，入表金额迅速增加（如图 6）。从行业分布来看，信息通信、软件和信息技术服务业、交通业、金融业等其数字化程度高、数据资源丰富的优势，在入表企业数量中占比相对更高，三大通信运营商的入表金额合计达 13.66 亿元，占全部上市公司

入表总金额的 44.38%¹¹。从入表的科目来看，以三大通信运营商为代表的多数企业，因数据资源多用于内部运营支撑与服务创新，符合无形资产定义，将其确认为无形资产；以海天瑞声为代表的数据服务商，其数据资源主要用于销售，因而将其确认为存货。从效果来看，在财报披露中展示数据资产的实际价值，可以增强投资者对企业数据储备和数据能力的量化认知，同时，推进数据资源入表过程中，需对数据资源的类型、内容和对应的应用场景进行盘点和分析，也可以进一步地帮助企业分析发掘和丰富数据应用方式。



来源：中国信息通信研究院

图 6 2024 年数据资源入表趋势

数据资产化帮助企业通过金融手段争取融资支持。根据不完全统计，截至 2025 年 3 月 31 日，国内非上市公司凭借入表的数据资源，

¹¹ 数据来源：《中国企业数据资产入表情况跟踪报告》2024 年年度报告[R].上海：高金智库数据资产课题调研组，2025：1-10.

已获得融资(包括银行授信、作价入股等形式)项目共 112 个，合计金额达 14.12 亿元。例如，平安证券主导发行的“平安一如皋第 1 期资产支持专项计划（数据资产）”于 2025 年 4 月在深圳证券交易所设立，规模达 1.3 亿元，该产品以如皋市交通、港口等领域多维度数据资产为底层质押，结合数据的具体应用场景、综合成本、收益和市场比较等维度构建估值模型，实现基于数据未来收益的结构化融资。

尽管上述企业的数据资产化探索已初显成效，但总体来看数据资产化探索仍面临诸多难题。一是数据资产权属界定模糊，不同主体对数据权利主张易产生纠纷。二是数据价值具有显著的波动性，不易精准估量，已形成的数据估值实践也面临行业评估标准不统一、跨企业估值结果缺乏可比性等具体挑战。三是从产业全局看，多数企业的数据资产全过程管理仍有环节上或能力上的欠缺。

五、“保安全”：数据智能紧密结合共筑安全堡垒

数据要素价值释放对产业升级与竞争力提升至关重要，但在人工智能跨越式发展的背景下，数据安全风险也随之而来，为数据要素的价值释放带来新挑战。因此，需要制度、技术、产业各方面共同发力，筑牢数据安全底线。2025 年，数据安全相关的制度要求进一步向垂直领域的精准治理深化，相关技术工具与人工智能结合得更加紧密，数据安全治理也正在从被动应对迈向主动治理的新阶段。

（一）制度进展：数据安全治理向 AI 领域延伸

在人工智能加速发展的环境下，针对生成式人工智能的数据安全治理要求开始强化，进一步完善数据安全治理框架。随着数据要素市

场化配置加速推进，我国逐步构建起涵盖数据安全、网络安全、个人信息保护等在内的数据安全基础法律框架。以《网络安全法》《数据安全法》《个人信息保护法》为核心，辅以多项部门规章和地方性法规，通过高位阶法律确立基本原则和责任边界，强化制度刚性，初步形成了覆盖数据全生命周期的制度体系。与此同时，针对人工智能这一垂直领域发展带来的新型风险，相关针对性立法稳步推进。2024年9月公布的《网络数据安全管理条例》设置生成式人工智能专条，明确加强对训练数据和训练数据处理活动的安全管理，完善生成式人工智能数据安全制度。目前，更具系统性和前瞻性的人工智能相关法律法规正在制定中，将为数据要素的健康发展提供法治保障。

分类分级监管成主流范式，“精准治理”让AI风险防控与创新发展同频共振。《生成式人工智能服务管理暂行办法》明确对生成式人工智能服务实行包容审慎和分类分级监管，由国家有关主管部门针对生成式人工智能技术特点及其在有关行业和服务应用，完善与创新发展相适应的科学监管方式，制定相应的分类分级监管规则或者指引。基于风险等级的分类分级监管正逐渐成为数据安全治理的主流范式，这一模式强调深入到具体领域场景中，根据不同类型的潜在风险程度，采取差异化的监管策略，实现精准施策与动态调整的有机统一。例如国家互联网信息办公室关于发布生成式人工智能服务已备案信息的公告中，将生成式人工智能服务分为“国家网信办备案”和“地方网信办登记”两类，备案面向具有舆论属性或社会动员能力的服务，登记面向通过API接口调用已备案模型能力的应用或功能，既

对高风险服务实施国家层面严格备案审查，又对低风险功能采用地方登记简化流程。截止 2025 年 6 月，已完成 439 款备案和 233 款登记¹²。这种治理方式不仅提升了监管的科学性和针对性，也为企业数据要素创新发展提供了合规空间，有效平衡安全保障与发展活力的关系，为数字经济高质量发展注入新动能。

（二）技术热点：数据安全技术愈加精准智能

人工智能技术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，传统依赖人工判断的方式已难以应对海量、多样的非结构化数据。当前，基于大模型等人工智能技术的多模态识别模型，可实现非结构化数据的自动化标签标注与敏感信息定位，降低人工标注成本，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。数据分类分级单点技术能力的智能化跃迁，为数据安全治理奠定了坚实的技术基础。

人工智能原生平台驱动数据安全运营形成智能防御体系。在数字经济快速发展和数据要素加速流通的背景下，传统依赖人工干预、规则匹配的数据安全运营模式已难以应对日益复杂和高频的安全威胁。人工智能技术的原生融合正在重塑数据安全运营范式，推动从“被动响应”向“智能防御”转型。人工智能驱动的安全运营平台能够整合来自威胁情报、网络日志、用户行为、终端设备等多源异构数据，利用大模型等技术构建企业数据的全局风险画像。在此基础上，系统可

¹² 《国家互联网信息办公室关于发布生成式人工智能服务已备案信息的公告》

实现对潜在攻击路径的预测判断，对异常行为的实时识别与溯源追踪，从而大幅提升威胁发现效率与处置精准度。此外，人工智能还能不断从历史事件中学习并优化策略，形成动态演进的防御机制，为数据安全提供持续、智能、闭环的保障能力。

（三）产业热点：安全治理从合规达标转向主动治理

随着数智技术在产业中的深度渗透，数据安全治理的战略地位正发生根本性转变。过去，企业往往将数据安全视为合规成本和风险控制的附属项，资源投入有限且缺乏系统规划。如今，在数据要素市场化配置改革加速推进、智能技术广泛应用的大背景下，数据安全已成为保障业务连续性、提升竞争力的关键能力。越来越多的企业开始将数据安全体系建设纳入战略顶层设计，视其为驱动创新、构建信任生态和实现可持续发展的“价值引擎”。通过有效的数据安全治理，企业能够保护所持有数据的安全，避免因风险事件导致损失，维护良好的品牌形象。同时，安全可靠的环境还能吸引更多客户和合作伙伴，拓展业务领域。

数据安全生态关系从企业“孤军奋战”向政产学研“全域联动”转变。产业发展初期，数据安全治理呈现出“孤军奋战”的状态，企业、安全厂商、监管机构等各方缺乏有效的沟通协作，安全信息难以共享，安全措施难以协同，导致安全防护效果不佳，难以应对日益复杂的网络攻击。随着产业生态关系日益复杂，数据安全治理正朝着“全域联动”的方向重构。企业与安全厂商加强合作，共同研发和部署先进的安全技术和解决方案；监管机构通过制定统一的安全标准和规范，

引导企业加强安全治理；行业协会等组织则搭建交流平台，促进各方之间的信息共享和经验交流。

供应链安全成焦点，从局部管控拓展至全链条保障。在数据产业链逐步融合的背景下，供应链安全问题日益凸显，成为数据安全治理的重点领域。传统安全治理多聚焦于企业自身系统与数据，而对上下游合作伙伴、第三方服务提供商等环节关注不足，导致攻击者常通过合作、供应等过程中的薄弱环节实施渗透，造成重大损失。当前，产业倾向于建立覆盖整个供应链的安全管理体系，推动安全管控从单一节点向全链条延伸。企业逐步加强对各环节供应链的安全审查和评估，要求供应商遵守严格的安全标准和规范。同时，建立供应链安全监测和预警机制，实时掌握供应链中的安全动态，及时发现和处理潜在风险。

六、发展展望

2025 年是数据要素市场建设继续高歌猛进的一年。面向“供得出、流得动、用得好、保安全”的目标，围绕数据基础制度、数据基础设施、数据融合应用、数据产业生态的多角度先行先试火热开展，数据要素的资源供给、流通交易、场景应用、安全治理水平正在积极而稳步提升中。然而，随着探索实践的拓宽和深入，我们也逐渐看到当前围绕数据基础制度构建的相关政策举措与已有制度间的衔接还不够顺畅，技术与产业实践中不断涌现新热点、新方向但突破性、规模化的成效尚不明显，各主体难以抽丝剥茧梳理出具体有效的发展路径，一系列新的、更加细节琐碎但又无可参考经验的问题和挑战还在

不断暴露出来。我们越来越深刻意识到数据事业是创新的事业、长期的事业，在推动新要素与传统要素融合、新机制与已有制度衔接、新模式对原有方案改良的过程中，如何保持定力、坚持长期投入、驰而不息，统筹发展与安全、持续激发各方参与的内生动力，将是未来发展的持续议题。

（一）容错免责机制设计需成为改革关键对象

我国数据要素发展，需要进行既符合数据要素的特征规律又能与现有制度有序衔接的制度和机制设计。工业生产的长期发展，形成了以标准化批量复制为目标的线性分工模式，与以数据要素为基础的、打破时空边界的、跨域融合的场景化运行模式有着本质区别。数据要素打破信息差、促进发展创新的潜力与泄露信息、引致安全风险的隐患同时存在，即使从国际视野看，平衡发展与安全也是探索过程中的共性难题。在建设全国统一大市场、建设数字中国的战略引导下，数据要素领域的改革应优先确定促发展的总基调，营造干事创业的良好探索氛围。在进行顶层设计和统筹规范的同时，要着重鼓励创新探索与实践完善，充分鼓励和发挥各行业、各地区、各类主体的积极性、主动性和创造性，使广大参与主体能安心“小步快跑、迭代优化”，不断探索符合数据特性的产业发展模式。

具体而言，要加快细化容错免责相关实施办法。当前，数据领域各项政策虽普遍提及“建立健全鼓励创新、包容创新的容错纠错机制”，但仍为方向性的表述，具体的实施细则尚未明确，对消除各类主体顾虑的作用不明显。针对当前各界在数据要素发展过程中的顾虑，要认

真落实党中央、国务院各项会议精神，认真落实习近平总书记强调的“三个区分开来”，即把干部在推进改革中因缺乏经验、先行先试出现的失误和错误，同明知故犯的违纪违法行为区分开来；把上级尚无明确限制的探索性试验中的失误和错误，同上级明令禁止后依然我行我素的违纪违法行为区分开来；把为推动发展的无意过失，同为谋取私利的违纪违法行为区分开来。为担当者担当、为干事者撑腰。

为此，**一是要明确规则底线**，探索制定数据产业发展相关负面清单，明确不可参与的数据范围和业务形式，定期评估、排查、清理各类显性和隐性壁垒，推动“非禁即入”普遍落实。**二是要宽容失败**，加快研究制定动机态度、客观条件、程序方法、性质程度、后果影响、挽回损失、整改成效等方面的判定条件。可在实施办法中明确，对危害后果轻微并及时改正的，或当事人有证据足以证明没有主观过错的，不予追责或处罚。可明确由于创新出现的首次不合规在限定时间内整改完毕，依法不予处罚。**三是要明晰责任边界**，由于任何技术和管理手段都不能确保万无一失，针对各类主体采取合规措施后仍无法应对风险事件追责的顾虑，要本着“尽职免责或免罚”的原则，建立数据技术或环境达标免责认定相关标准，健全合规后免罚、推定无过错等机制。**此外，各类主体在创新探索中也应自律**，充分考虑本行业、本地区、本企业数据发展基础和阶段，理顺发展思路，稳妥选择发展路径，避免盲目跟风、追求速成的思维，从而推动数据深度赋能各业务环节、各行业领域，为经济社会的高质量发展注入蓬勃生机。

（二）三大引导方向或将在短期内快速推进

我国数据要素市场化配置改革逐步进入以体系化建设和融合创新为特征的高质量平稳发展期。在国家数据局的统筹引领下，数据基础设施建设、公共数据资源开发利用、数据产业主体培育三大方向，或将作为相互支撑、协同共进的战略支点，在短期内快速推进，为构建安全可信、繁荣开放的全国一体化数据市场奠定良好基础。

数据基础设施将作为数字中国建设的重要底座加速建设。数据基础设施建设是一项开创性、系统性工程，很难一蹴而成。当前 18 个城市的先行先试，将是不断发现问题、探索解决问题、优化完善建设方案并积累建设经验的必然过程。一方面，先行先试将对各类技术方案进行充分验证，并推动各类关键技术方案加强突破升级，向更高适配性、更灵活易部署、更低成本的方向融合创新，以满足海量数据的高实时、高并发、高准确性、强安全的要求；另一方面，“硬建设”陆续就绪的同时，也将带动一系列“软建设”的探索推进，数据流通合规准入、数据价值收益分配等支撑基础设施可持续运行相关机制建立将同步推进，数据标注、产品开发、场景挖掘、供需撮合等提供配套专业服务的相关产业也将加快培育与激励，激活数据生态。

公共数据资源开发利用将从制度层面解决数据“供不出”难题。

《政务数据共享条例》已在法规层面为打破政府部门“数据孤岛”提供了高级别制度保障。公共数据资源开发利用“1+3”政策体系正在加速构建全国公共数据资源“一本账”，推动公共数据以多样化形式向社会供给。预计未来一段时间，《关于加快公共数据资源开发利用

的意见》所要求的“公共数据资源授权运营情况披露机制”“公开公共数据产品和服务能力清单”等将逐步落地，授权运营形成的公共数据产品和服务可在相关网站或平台查询，在公众监督下被市场主体再开发再利用。

政策“组合拳”将有效培育专业化、规模化的数据产业主体。部分地方已明确提出数据企业数、数据产品数等量化目标，并配套了财税政策、算力补贴、金融产品、产业基金等保障措施，着力培育壮大数据产业链。举办赛会、组织试点、案例推广等引导方式将吸引和带动不同类型企业增强数据意识，启发、带动潜力企业提升数据能力，调动、激活数据开发利用积极性。同时，政府部门或将优化采购工作流程，变项目采购模式为数据产品和服务采购模式，增加采购相关科目，构建社会数据“统采共用、分采统用”等机制，以便带头采购数据产品和服务，撬动数据产业发展。

（三） 夯实企业能力仍是发挥市场主导作用的基础

企业是产业的基石，是数据要素市场中供给、需求和创新的核心主体。然而，微观层面的企业数据能力建设与宏观层面的政策热潮仍不完全适应，市场主导作用的发挥受阻。数据要素市场的决定性作用理应通过企业间的供需互动来发挥，但企业数据管理、应用等能力的系统性缺失导致数据内部应用、对外供给和引入外部数据深度应用等行为均不够充分。除少数数据基础牢、经营实力强的大型企业投入专门力量进行深度挖掘数据价值外，更广泛的企业群体由于缺乏专业素养和技术能力，不仅不知道哪些业务可以被数据改善优化，数据本身

反而因存储和管理成为一种拖累的成本而非创造价值的资产，更难产生引入高价值外部数据的需求。相应地，市场上的数据产品常停留于浅层次应用水平，成熟的数据产品多集中于企业征信、信息核验等领域。这类数据产品由于数据来源相对单一、技术门槛不高，已然陷入竞争激烈、利润微薄的“红海”。而充分体现数据要素乘数效应的、基于跨行业、跨部门、跨地域数据融合的高附加值应用场景，则因企业数据能力短板而难以涌现。因此，夯实企业的数据能力，是激活市场内生动力、引导市场发挥资源配置决定性作用的关键基础。

天然具有“数据基因”的数据龙头企业能力强，可致力于实现自身数据能力的战略性扩散。数据龙头企业以持有海量高价值数据、掌握流量渠道、汇聚大批高素质数据人才的互联网平台、运营商和金融科技等企业为代表，各行业内扮演类似角色的企业亦可作为本行业的数据龙头。作为产业链的“链主”，龙头企业不能仅仅满足于自身业务的闭环，更应扩散自身的数据能力，帮助产业链内更多企业建立和夯实能力，完善产业生态体系。一方面，龙头企业可将内部成熟的数据专业团队与业务板块剥离，自主孵化出独立的专业数据企业，为市场提供高质量的数据治理、数据分析或模型训练服务，成为数据产业专业服务体系的中坚力量。另一方面，龙头企业可以向产业链上下游及合作伙伴开放其数据处理平台、算法工具库和应用开发接口，导入生态资源，降低中小企业使用先进技术、吸收成熟经验和方法论的门槛。

相比数据龙头企业，传统企业和中小企业的首要任务不是追求前

沿创新，而是扎实建立和巩固数据能力。这些企业由于人员、技术、资金、认知等的限制，普遍存在数字化转型基础薄弱、“不敢转、不会转”的困境。一方面，这些企业可积极拥抱和参与龙头企业搭建的合作平台，对接高价值的生态资源、吸收借鉴其提供的开放能力，但另一方面，仍必须明确，企业数据能力提升的起点在内部。企业需要结合自身业务发展目标，构建基础的数据管理流程，在具体的业务场景中从小范围数据应用开始尝试，逐步建立起专业的数据管理、应用等能力。只有当企业能够规范地采集、加工、使用数据，才能为开发数据产品或有效引入外部数据奠定基础，这也是广大企业参与全国一体化数据市场的前序环节。根据产业分工和市场定位的不同，部分企业可致力于成为深耕行业细分领域的“专精特新”型企业，将自己放大为服务整个行业的创新动能。通过长期聚焦某一专业领域，深刻洞察行业痛点，不断夯实独特的数据能力，在实践中逐步形成有效的数据产品和解决方案，从而支撑产业链高效运转。

中国信息通信研究院云计算与大数据研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62302928

传真：010-62303980

网址：www.caict.ac.cn

