

全国数据资源调查报告(2025年)

NATIONAL DATA RESOURCE SURVEY REPORT

全国数据资源统计调查工作组

版权声明

本报告版权属于全国数据资源统计调查工作组，并受法律保护。
转载、摘编或利用其他方式使用本报告文字或者观点的，应注明“来源：《全国数据资源调查报告（2025年）》”。违反上述声明者，将追究其相关法律责任。

专家委员会

(排名不分先后)

顾问

邬贺铨 江小涓 梅宏 刘韵洁 尹浩 张平文 李骏 孙胜利
高新民 许宪春

委员

江明涛 姜奇平 廖凯 汪玉凯 李晓东 黄罡 张向宏 何强
吴志刚 蔡跃洲 王建民 欧阳日辉 王晨 赵刚 孙毅 胡良霖
林伟 王钺 冯海红 杨翰方 郑爱军

工作组

国家工业信息安全发展研究中心

复旦大学上海数学中心

前言

INTRODUCTION

习近平总书记指出，当今世界正在加速迈入智能时代，数据的基础资源作用和创新引擎作用日渐显现。党中央、国务院高度重视数据工作，就健全数据要素基础制度、深化数据资源开发利用、促进实体经济和数字经济深度融合等方面作出全面部署。国家数据局成立以来，贯彻落实党中央、国务院决策部署，坚持以数据要素市场化配置改革为主线，围绕数据基础制度、数据基础设施、数据融合应用和场景建设、全国一体化数据市场建设、数据产业培育5个方面打出一套“组合拳”，夯实数据领域核心技术攻关、数据标准化、数字人才培养3个基础，突出数据赋能人工智能创新发展1个重点，探索形成了“531”的工作体系，持续推动数据“供得出、流得动、用得好、保安全”，促进数据要素全面赋能经济社会发展。

贯彻落实《中华人民共和国国民经济和社会发展第十五个五年规划纲要》提出“构建国家数据资源体系，健全数据资源统计调查制度，建立全国数据资源‘一本账’”的工作部署，国家数据局修订了《全国数据资源统计调查制度》。在此基础上，组织开展2025年度全国数据资源统计调查工作，调查有效样本数量4.25万个，覆盖除国际组织外所有国民经济行业门类。在保持数据供给、流通与开发利用等指标体系连续性的基础上，本年度调查实现了双向扩容：一是向“新主体”拓展，将地方数据集团、可信数据空间等纳入调查范围；二是向“新热点”拓展，将词元(Token)调用、高质量数据集流通交易、企业数据要素投入、数据增值等产业发展热点作为调查研究重点，持续提升对我国数据资源全景刻画、结构性分析和前瞻性研判能力，形成《全国数据资源调查报告(2025年)》。

调查结果显示，我国数据要素市场化价值化进程显著提速，数据资源供给体系、数据流通体系和数据开发利用体系协同演进，正由数据资源规模扩张向数据要素价值释放跃升，呈现以下显著特征：

一是数据资源规模保持高速增长态势。2025年，全国年度数据生产总量¹达52.26泽字节（ZB），同比增长27.28%，增速较上年提升2.28个百分点。我国数据生产总量占全球约27.44%。**从主体看**，企业成为数据生产的主力，数据生产增量约九成来自企业数据，产业数智化转型成效明显。**从行业和领域看**，工业制造、交通物流、软件和信息技术服务业等行业数据生产量增幅较大，同比分别增长1.27泽字节（ZB）、1.22泽字节（ZB）、0.92泽字节（ZB），有效发挥“领头羊”“压舱石”作用。具身智能、低空经济等新兴领域数据生产量高速增长，增速分别为477.78%、75%。全国数据存储总量达2.53泽字节（ZB），同比增长21.05%。结构化数据存储量为0.56泽字节（ZB），同比增长43.59%，占数据存储总量的22.13%，数据质量持续提升，加速向可计算、可分析形态转化。

二是算力基础设施建设稳步推进。“东数西算”工程全面实施，全国一体化算力网加快构建，智能算力供给水平稳步提升。截至2025年底，全国智能算力规模达159万PFLOPS（FP16），通用算力向智能算力的代际更替加速，成为支撑人工智能发展的关键基础设施。智算资源集聚优势逐步显现，八大国家算力枢纽（含十大集群）智算规模占全国智算规模的80%以上。

1. 全国年度数据生产总量指2025年度全国数据创建、复制、捕获和消费产生的数据总量。

三是数据资源开发利用更加高效。随着“数据要素×”行动、公共数据“跑起来”示范场景建设、国有企业数据效能提升行动、国家数据基础设施建设先行先试、可信数据空间发展行动计划等纵深推进，数据融合应用和场景建设不断深化，数据要素价值加速释放。**公共数据资源开发利用成效显著。**公共数据用于共享、开放、授权运营的数据量均快速增长，申请共享的数据集数量同比增长近30%，公共数据开放数据量和授权运营数据量同比分别增长31.71%、53.96%。公共数据带动各行业数据加快融合应用，场景覆盖产业发展、教育科技、医疗健康、民生服务、基层治理等领域。**企业用数活力加速释放。**2025年，企业数据技术投入同比增长17.37%。企业数据产品和服务数量同比增长29.29%，交易额同比增长39.8%，数据产品和服务从数字化副产品向赋能业务增长的驱动引擎转变。

四是数据流通交易效果初显。全国一体化数据市场建设加快推进，市场活力进一步激发，数据要素价值加速释放，为高质量数据付费的市场共识正在形成。**数据流通活跃度不断提高。**2025年，全国数据跨境流通总量为142.34艾字节（EB），同比增长14.88%。数据跨省流通总量为2949.12艾字节（EB），同比增长19.01%，广东、浙江、江苏、山东、河南等经济大省数据跨省流通量位居前五。企业数据流通总量为1935.36艾字节（EB），同比增长25.17%。头部平台企业和中央企业²的数据流通规模均值是其他企业的140多倍，数据流通枢纽功能持续强化。**数据付费意愿不断提升。**样本企业中，11.65%的企业购买过数据，购买数据费用同比增长22.36%。头部平台企业和中央企业购买数据费用均值是其他企业的60倍。金融、软件和信息技术服务业等行业购买过数据的企业比例均超30%，大幅高于行业平均水平。

五是数据赋能人工智能迈入规模化应用新阶段。人工智能发展从通用大模型到行业垂域模型，再到代理型人工智能形态演进，对数据的需求

2. 本报告中央企业有效样本数量为113家。

也从基础语料到行业高质量数据集拓展。调查表明，2025年，用于人工智能训练和推理的数据总量为199.48艾字节（EB），同比增长42.86%，推理数据量达101.34艾字节（EB），首超训练数据量。高质量数据集数量超11万个，规模超908拍字节（PB），同比分别增长61.13%和142.58%。全年词元（Token）调用量约21100万亿，词元（Token）正成为人工智能的新度量衡。

同时也要看到，当前我国数据市场建设还处于起步阶段，数据价值释放仍有较大潜力。2026年是“十五五”开局之年，也是“数据要素价值释放年”，随着数据基础制度加快健全完善、数据基础设施规模化部署和应用能力不断增强、全国一体化数据市场建设持续推进，数据要素在更大范围、更深层次参与资源配置的条件正加速形成。展望未来，我国数据资源规模优势将加快向价值优势转化，数据要素将在赋能经济社会发展和治理能力提升、促进生产方式深层次变革和生产力革命性跃迁中发挥更加重要的基础性、战略性作用。

报告不足之处在所难免，敬请各位专家和广大读者指正。

目录

CONTENTS

第一章 重点指标	01
一、数据资源供给主要指标	02
二、数据要素流通主要指标	03
三、数据资源开发利用主要指标	04
第二章 调查分析	05
一、数据资源供给规模持续扩大	06
(一) 数据生产规模保持高速增长	06
(二) 存储数据可用性提高	07
(三) 智能算力持续保持高速扩张	08
二、数据流通交易效果初显	08
(一) 数据流通活跃度不断提高	08
(二) 数据付费意愿不断提升	09

三、数据资源开发利用更加高效	09
(一) 数据赋能人工智能创新发展	09
(二) 公共数据资源开发利用不断深化	11
(三) 企业用数活力加速释放	11
四、数据加快赋能千行百业	12
(一) 农业数据助力智慧农业建设	12
(二) 工业供数用数动力逐渐提升	12
(三) 服务业数据流通利用成效显著	13
(四) 科学数据引领创新应用与转化	13
第三章 发展趋势	15
一、人工智能将变革数据生产利用方式	16
二、词元成为数据价值释放的新路径	16
三、数据资源流通效率加快提升	16
四、数据产业专业化分工体系不断深化	16
附录: 调查方法	17

第一章 重点指标

一、数据资源供给主要指标

全国数据生产、存储规模保持高速增长，活跃数据量明显提升，通用算力向智能算力的代际更替加速，数据生产、存储和开发利用的“剪刀差”开始缩窄。

主要指标		2024年 新增值	2025年 新增值	增长率	累计值 (截至2024年)	累计值 (截至2025年)	增长率
数据生产	全国年度数据生产总量 (ZB)	41.06	52.26	27.28%	—	—	—
	— 物联网感知数据生产量 (ZB)	21.17	25.34	19.70%	—	—	—
	— 系统/软件/AI数据生产量 (ZB)	19.89	26.92	35.34%	—	—	—
	人均数据生产量 (TB/人)	31.31	39.94	27.56%	—	—	—
	企业数据生产总量 (ZB)	23.94	33.54	40.10%	—	—	—
数据存储	数据存储总空间 (ZB)	—	—	—	3.42	3.94	15.20%
	数据存储总量 (ZB)	—	—	—	2.09	2.53	21.05%
	云存储数据量 (ZB)	—	—	—	0.64	0.98	53.13%
	结构化数据存储量 (ZB)	—	—	—	0.39	0.56	43.59%
	活跃数据量 (ZB)	—	—	—	1.3	1.67	28.46%
算力支撑	智能算力规模 (PFLOPS) (FP16)	—	—	—	—	159万	—

二、数据要素流通主要指标

跨境、跨省数据流通持续提速，头部平台企业和中央企业发挥数据资源汇聚、流通和开发利用的枢纽作用，为高质量数据付费的市场共识正在形成。

主要指标		2024年 新增值	2025年 新增值	增长率	
数据流通	数据跨境流通总量 (EB)	123.90	142.34	14.88%	
	—企业数据跨境流通总量 (EB)	75.77	88.06	16.22%	
	数据跨省流通总量 (EB)	2478.08	2949.12	19.01%	
	公共数据流通总量 (EB)	—	440.32	—	
	企业数据流通总量 (EB)	1546.24	1935.36	25.17%	
	数据付费	购买过数据的企业比例 (%)	10.58	11.65	10.11%
	企业购买数据费用增速 (%)	—	22.36	—	

三、数据资源开发利用主要指标

数据资源支撑人工智能跃迁式发展，词元（Token）调用量呈现指数级增长。数据资源开发利用不断深入，公共数据供给方式从“共享为主”向“共享、开放、授权运营协同推进”转变，企业用数活力不断提升，加快释放数据要素价值。

主要指标		2024年 新增值	2025年 新增值	增长率	累计值 (截至2024年)	累计值 (截至2025年)	增长率	
数据开发利用	人工智能	用于人工智能的数据总量 (EB)	—	—	—	139.63	199.48	42.86%
		高质量数据集数量 (万个)	—	—	—	7.10	11.44	61.13%
		高质量数据集总数据量 (PB)	—	—	—	374.43	908.30	142.58%
		年度大模型词元调用数量 (万亿个)	—	21100	—	—	—	—
		头部平台企业人工智能开发训练 中使用企业外部数据比例 (%)	54.25	74.36	37.07%	—	—	—
	公共数据	政务数据共享枢纽调用次数 (亿次)	—	—	—	>5400	>5500	—
		公共数据开放数据量 (PB)	—	—	—	28.73	37.84	31.71%
		公共数据授权运营数据量 (PB)	—	—	—	4.93	7.59	53.96%
	企业数据	数据技术投入增速 (%)	—	17.37	—	—	—	—
		实现数据增值的企业比例 (%)	—	—	—	—	11.14	—
		数据资产入表金额增速 (%)	—	34.63	—	—	—	—
		数据产品/服务平均数量 (个)	—	—	29.29%	—	—	—
		数据产品/服务交易额增速 (%)	—	39.80	—	—	—	—

第二章 调查分析

一、数据资源供给规模持续扩大

（一）数据生产规模保持高速增长

2025年，全国年度数据生产总量达52.26泽字节（ZB），同比增长27.28%，增速较上年提升2.28个百分点。我国数据生产总量占全球约27.44%。人均数据生产量约为39.94太字节（TB），同比增长27.56%。

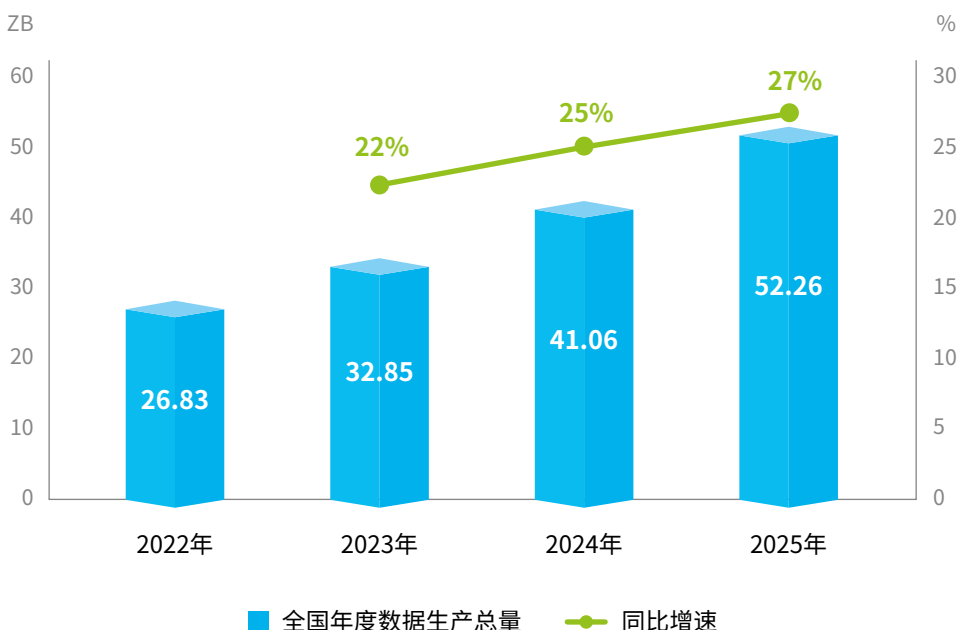


图1 2022—2025年全国年度数据生产总量

从结构看，数据生产正由感知驱动向计算驱动转变，系统软件、人工智能产生的数据量为26.92泽字节（ZB），首次超过摄像头、传感器等物联感知数据量25.34泽字节（ZB），用于人工智能训练、开发和应用的数据增长显著。

从主体看，企业成为数据生产的主力。2025年，企业数据生产总量为33.54泽字节（ZB），占数据生产总量比重从58.30%提升到64.18%，贡献九成数据生产增量，产业数智化转型成效明显。

从行业和领域看，工业制造、交通物流、软件和信息技术服务业等行业数据生产量增幅较大，分别增长1.27泽字节（ZB）、1.22泽字节（ZB）、0.92泽字节（ZB），有效发挥“领头羊”“压舱石”作用。具身智能、低空经济等新兴领域数据生产量高速增长，增速分别为477.78%、75%。

（二）存储数据可用性提高

2025年，全国数据存储总量为2.53泽字节（ZB），同比增长21.05%。存储空间为3.94泽字节（ZB），存储空间利用率为64.21%，较上年提升3个百分点。

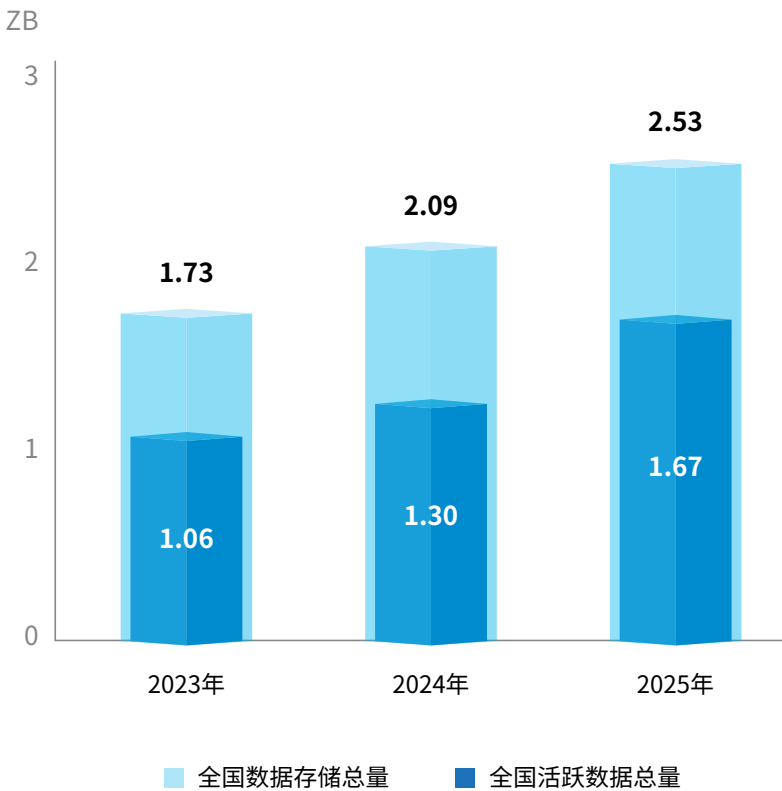


图2 2023—2025年全国数据存储总量及活跃数据总量

从存储结构看，云存储总量为0.98泽字节（ZB），同比增长53.13%，占数据存储总量的38.74%，有效支撑了业务协同和数据交互。结构化数据存储量为0.56泽字节（ZB），同比增长43.59%，占数据存储总量的22.13%，数据质量持续提升，加速向可计算、可分析形态转化。

从利用效率看，人工智能的普及应用极大促进了存储数据的盘活利用，全国活跃数据总量为1.67泽字节（ZB），同比增长28.46%，增速超过数据存储总量增速。在“人工智能+”行动的政策推动下，中央企业平均活跃数据量超56拍字节（PB），同比增长41.98%，示范带动作用明显。

（三）智能算力持续保持高速扩张

截至2025年底，全国智能算力规模为159万PFLOPS（FP16），通用算力向智能算力的代际更替加速，成为支撑人工智能发展的关键基础设施。

从空间布局看，智算资源集聚优势逐步显现，八大国家算力枢纽（含十大集群）智算规模占全国智能算力的80%以上。贵州、内蒙古和河北加快打造算力高地，占全国智算规模超六成，全国一体化算力网建设加快。

从建设主体看，中央企业和平台企业加大对智能算力的战略性投资，智算资源成为人工智能时代衡量企业竞争力的重要指标。中央企业自建算力中，智能算力占比近80%，智能算力规模同比增长65.99%。头部平台企业智算建设投入超千亿。

二、数据流通交易效果初显

（一）数据流通活跃度不断提高

2025年，数据流通交易政策文件加快落地，全国一体化数据市场建设持续推进，跨境、跨省数据流通更加活跃。

数据跨境流通稳步增长。2025年，全国数据跨境流通总量为142.34艾字节（EB），同比增长14.88%。其中，企业数据跨境流通量为88.06艾字节（EB），同比增长16.22%，数据领域国际合作更加深化。

数据跨省流通更加活跃。数据跨省流通量为2949.12艾字节（EB），同比增长19.01%。广东、浙江、江苏、山东、河南等经济大省数据跨省流通量位居前五，数字经济辐射效应显著。

公共数据流通带动作用不断增强。公共数据流通量为440.32艾字节（EB），其中超90%的数据在政府和企业之间流通，公共数据和企业数据融合应用不断加快。地方数据集团数据流通量中，66.1%为公共数据，数据集团推动公共数据市场化的作用初显。

企业数据流通规模快速增长。企业数据流通量为1935.36艾字节（EB），同比增长25.17%。头部平台企业数据流通规模是其他企业的140多倍，数据流通枢纽功能持续强化。

（二）数据付费意愿不断提升

2025年，样本企业中，11.65%的企业购买过数据，购买数据费用同比增长22.36%，为高质量数据付费的市场共识正在形成。

从主体看，人工智能模型企业购数需求旺盛，头部平台企业在人工智能开发训练过程中使用企业外部数据占比达74.36%，购买数据费用均值是其他企业的60倍，购买数据量均值是其他企业的115倍，人工智能进一步激发企业用数活力。

从行业看，金融、软件和信息技术服务业等行业购买数据企业比例均超30%，大幅高于行业平均水平。

从区域看，北京、上海积极打造人工智能发展高地，为高质量数据付费内生动力强劲，购数企业比例均超30%，大幅高于全国平均水平。

三、数据资源开发利用更加高效

（一）数据赋能人工智能创新发展

2025年，用于人工智能训练和推理的数据总量为199.48艾字节（EB），同比增长42.86%，推理数据量达101.34艾字节（EB），首超训练数据量，人工智能加速规模化落地应用。

高质量数据集加快建设。2025年，全国高质量数据集数量超11万个，同比增长61.13%。高质量数据集数据量超908拍字节（PB），同比增长142.58%。

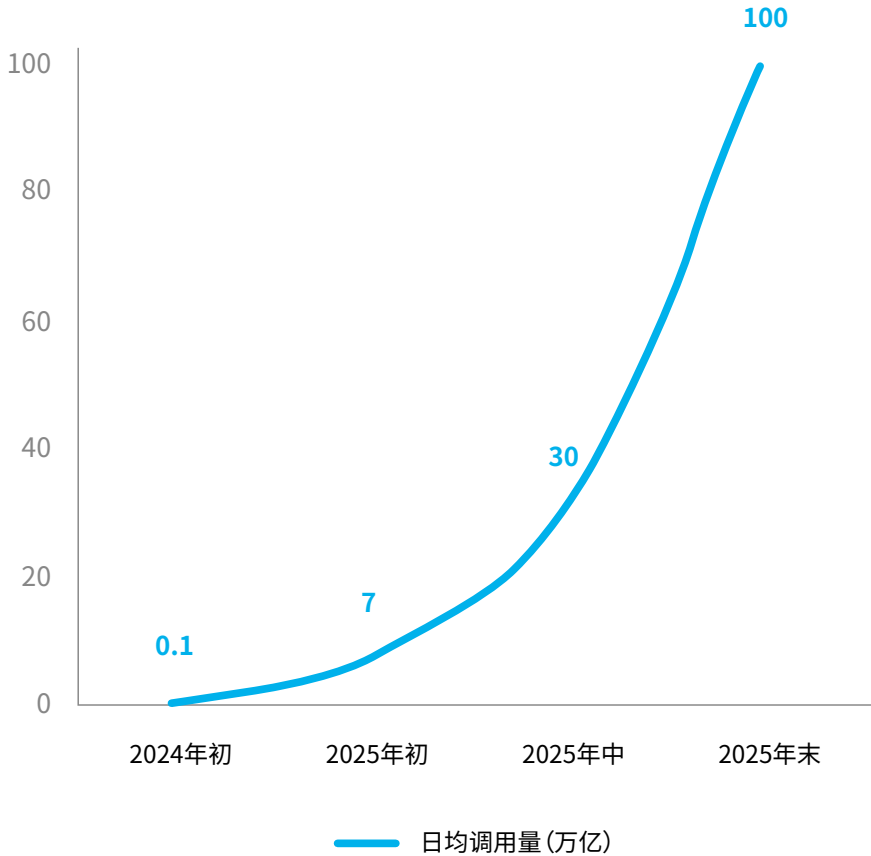


图3 2024—2025年日均词元 (Token) 调用量变化趋势

词元 (Token) 调用量指数级增长。2025年，全国日均词元 (Token) 调用量从年初的超万亿增长到年末的100万亿，智能经济的澎湃活力加速释放。全年词元 (Token) 调用量约21100万亿，“数据供给—价值释放”的良性循环初现端倪。

（二）公共数据资源开发利用不断深化

公共数据资源开发利用“1+3”政策体系落地见效，公共数据供给方式从“共享为主”向“共享、开放、授权运营协同推进”转变。

政务数据共享渠道更加畅通。全国一体化政务数据共享枢纽累计支撑调用超5500亿次，申请共享的数据集数量同比增长近30%，有力支撑“高效办成一件事”。

公共数据开放力度持续加大。全国公共数据开放数据量同比增长31.71%，省级公共数据开放数据量同比增长39.74%，为产业发展和公共治理提供有力支撑。

公共数据授权运营带动各类数据融合利用。全国公共数据授权运营数据量同比增长53.96%，省级公共数据授权运营数据量同比增长54.17%，授权运营的数据产品和服务数达1.02万个，应用场景覆盖产业发展、教育科技、医疗健康等多个领域。

（三）企业用数活力加速释放

随着“数据要素×”行动、国有企业数据效能提升行动、可信数据空间发展行动等纵深推进，企业加大新场景开发和推广力度，企业数据从服务内部“智改数转”，加快向对外输出增值服务延伸。

表1 企业数据开发利用情况

指标	头部平台企业	中央企业	数据企业	样本企业
数据开发率 (%)	53.51	20.36	34.36	28.46
购数企业比例 (%)	88.89	84.95	30.23	11.65
数据技术投入增速 (%)	25.79	24.49	22.11	17.37
实现数据增值的企业比例 (%)	25.00	26.17	18.25	11.14

企业持续加大数据技术投入。2025年，企业数据技术投入同比增长17.37%。从主体看，头部平台企业和中央企业数据技术投入同比分别增长25.79%、24.49%。

企业加大数据产品和服务供给。2025年，企业数据开发率³为28.46%，企业数据正成为产品升级、模式创新、价值创造的核心战略资源。企业数据产品和服务数量同比增长29.29%，交易额同比增长39.8%，数据产品和服务从数字化副产品向赋能业务增长的驱动引擎转变。可信数据空间试点机构数据开放量同比增长118.2%，孵化7000个数据产品与服务，在数据开放机制、数据产品化、商业运营等方面探索新模式。

企业数据价值加快释放。2025年，利用数据获得直接收益的企业占比为11.14%。企业数据资产入表金额同比增长34.63%，数据要素复用增效和融合创新不断提升。

四、数据加快赋能千行百业

（一）农业数据助力智慧农业建设

2025年，农业数据生产量为1.49泽字节（ZB），占企业数据生产总量的4.44%。在场景方面，农业领域的公共数据开放量同比增长42%，遥感、气象等领域公共数据助力精准农业生产。在技术投入方面，农业企业数据技术投入同比增长26%，用于人工智能的数据量同比增长14.51%，“智慧农业”加速发展。

（二）工业供数用数动力逐渐提升

随着智能制造、数字化转型、工业互联网创新发展等工程深入推进，制造业“智改数转网联”成效显著。在数据生产方面，2025年，工业数据⁴生产量为8.53泽字节（ZB），占企业数据生产总量的25.43%。工业设备数据占工业数据比例超三

3. 数据开发率=经加工、开发和复用的数据规模占存储数据规模的比重。

4. 按照国家统计局《国民经济行业分类（GB/T 4754-2017）》，第二产业包括工业和建筑业，其中工业包括采矿业、制造业、电力/热力/燃气及水生产和供应业。

成，机器人、电力设备、电子设备、化工设备等数据生产量占比较高，机器人、无人机等数据生产量同比分别增长51%、75%。在数据开发利用方面，制造业数据开发率为32.21%，领先全国平均水平3.75个百分点，数据利用效率优势明显。

(三) 服务业数据流通利用成效显著

2025年，服务业数据生产量为22.85泽字节（ZB），占企业数据生产总量的68.13%。服务业数据流通量（1.29泽字节）高于农业和工业，交通、金融行业数据活跃度（分别为76.19%、68.92%）领先其他行业，服务业数据跨区域、跨行业流通利用快速增长。

(四) 科学数据引领创新应用与转化

AI for Science推动科学数据创新涌现，加速向产业转化应用。在数据开发方面，13.43%的科学数据用于人工智能训练分析，同比增长50.32%，增速领先其他行业，科研用数需求高涨。在数据应用方面，90%的科学数据服务于前沿研究和重大工程项目，10%的科学数据向产业、民生领域转化应用，科学数据逐步赋能其他领域。

表2 行业数据分布

产业类型	国民经济行业门类	2025年数据生产量 (ZB)	企业数据流通			数据开发率 (%)	用于人工智能的数据比例		购数企业比例 (%)
			年度 (EB)	增速 (%)	公共数据占比 (%)		累计 (%)	增速 (%)	
一产	农、林、牧、渔业	1.49	32.08	16.01	16.03	11.52	1.57	14.51	3.21
二产	采矿业	0.76	18.38	21.98	15.73	27.71	1.67	17.54	3.94
	制造业	7.19	495.52	34.07	14.01	32.21	9.10	47.55	10.95
	电力、热力、燃气及水生产和供应业	0.58	26.68	23.86	18.04	22.70	4.58	48.16	13.25
	建筑业	0.67	40.96	21.63	15.15	10.40	1.53	34.33	8.18
三产	批发和零售业	1.50	93.71	18.36	17.62	16.35	3.51	32.21	5.44
	交通运输、仓储和邮政业	2.64	396.72	36.25	22.55	31.00	13.73	49.12	13.37
	住宿和餐饮业	0.06	3.30	16.49	10.96	12.39	0.18	13.63	2.15
	信息传输、软件和信息技术服务业	7.47	513.23	36.97	13.86	34.36	14.13	62.95	30.23
	金融业	2.89	37.88	29.22	27.82	32.61	12.16	46.41	30.45
	房地产业	0.88	41.28	14.95	9.79	10.66	1.22	12.11	1.36
	租赁和商务服务业	1.55	7.33	25.56	12.68	6.31	1.02	29.59	4.35
	科学研究和技术服务业	1.96	146.42	33.15	13.51	32.88	13.43	50.32	13.02
	水利、环境和公共设施管理业	1.19	7.44	11.98	13.13	10.19	6.07	29.81	7.87
	居民服务、修理和其他服务业	0.45	5.59	14.92	13.42	17.52	4.86	25.43	2.08
	教育	0.31	1.93	18.89	25.59	15.44	1.06	34.13	11.11
	卫生和社会工作	1.58	60.36	19.19	28.80	11.90	1.38	30.23	2.78
	文化、体育和娱乐业	0.37	1.92	16.58	10.84	11.09	3.93	33.12	6.27

注：按照国家统计局《国民经济行业分类（GB/T 4754-2017）》，第二产业包括工业和建筑业，其中工业包括采矿业、制造业、电力/热力/燃气及水生产和供应业。

第三章 发展趋势

一、人工智能将变革数据生产利用方式

预计2026年，全国年度数据生产总量将超过67泽字节（ZB）。以智能体（Agent）为代表的代理型人工智能快速发展，促进智能体互联网加速形成，海量智能体之间的协同交互将产生大规模机器间数据，传统以数据采集—存储—处理为主的业务链条，将逐步转向“实时感知—实时决策—持续优化”的业务闭环模式，数据结构和规模均将发生质变。同时，以具身智能为代表的物理AI加速发展，感知数据、控制数据与环境反馈数据将形成循环，推动数据从虚拟空间走向物理世界交互闭环，现实世界数据规模将出现新一轮跃升。

二、词元成为数据价值释放的新路径

随着人工智能应用深入，词元（Token）等新型计量方式，促进数据价值体系加速从“资源计量”向“使用计量”“效果计量”转变，数据流通交易从以数据资源为主的形态加速向“数据+算力+模型”一体化数据产品和服务模式演进。围绕数据使用形成的付费机制将逐步成熟，为高质量数据付费将成为业界共识，数据流通和交易规模将进一步提升，预计2026年全国活跃数据量将超过2泽字节（ZB）。

三、数据资源流通效率加快提升

2026年，企业数据流通量将保持两位数增长。随着数据要素基础制度逐步健全和数据基础设施应用落地，为数据流通利用提供有力保障。数据交易所（中心）、数据流通服务平台企业、数据商等数据流通服务机构在数据安全合规高效流通中发挥更大作用。智能体将参与数据发现、匹配、定价与交易过程，推动数据流通效率大幅提升，全国一体化数据市场加快形成。

四、数据产业专业化分工体系不断深化

随着数据市场的发展壮大，数据资源、数据技术、数据服务、数据应用、数据安全、数据基础设施六类数据企业，将沿数据全生命周期演化出完整的专业化分工产业链条，数据采集、清洗、标注、合成、开发、托管、运营等环节将涌现一批专业化服务商，催生数据标注、数据合规等新职业，促进数据供给、流通和利用的效率体系化提升，数据产业集聚效应将显现。

附录：调查方法

一、数据来源

面向14类调查对象开展数据资源调查工作，有效样本情况如表3所示。

表3 有效样本情况

序号	调查对象	调查主体		数量 (个)	
1	政府部门	国家行业主管部门		25	
2		省/自治区/直辖市及新疆生产建设兵团数据管理部门		32	
		计划单列市数据管理部门		5	
3	事业单位			2722	
4	科研机构			1259	
5	国家科学数据中心			10	
6	企业	中央企业	运营商	4	
7			中央企业 (不含运营商)		109
8		数据企业	平台企业		168
9			数据交易所 (中心)		41
10			地方数据集团		246
11			可信数据空间建设和运营商		826
12			其他数据企业		16818
13			其他企业		
14		行业协会商会			56

二、样本分布

报告有效样本覆盖全国31个省（自治区、直辖市）和新疆生产建设兵团，其中江苏、广东、山东、浙江、河北等省份样本企业较多。

表4 各省调查主体数量分布

地区	填报完成(家)	有效样本(家)	地区	填报完成(家)	有效样本(家)
北京	2035	1956	湖北	3759	3312
天津	526	336	湖南	1351	1082
河北	4078	3623	广东	4333	3937
山西	420	372	广西	285	261
内蒙古	705	624	海南	145	121
辽宁	1047	825	重庆	555	476
吉林	256	197	四川	2003	1706
黑龙江	975	864	贵州	381	302
上海	521	495	云南	425	385
江苏	5265	4723	西藏	10	7
浙江	4295	3853	陕西	946	639
安徽	2012	1841	甘肃	753	457
福建	2215	1933	青海	154	88
江西	2263	2093	宁夏	111	103
山东	4285	3887	新疆	256	211
河南	2017	1786	新疆兵团	14	10

三、统计方法

（一）样本处理

本次调查对样本数据采用双侧缩尾（截断）预处理，消除样本中极端异常值对统计结果的干扰，保留有效样本数据进入后续测算环节。该处理方式在保留样本整体分布特征的前提下，有效过滤了异常值带来的结果失真，避免填报误差、极端个例导致的估计偏误，确保后续测算的稳定性与可靠性。

（二）分层抽样

考虑到全国企业数量多且规模差异明显，分层抽样方法相对其他随机抽样方法更能兼顾总体中各类差异，保证测算结果准确性。本次调查采用分层抽样统计调查方法，将企业样本划分为不同层级，在每个层级中按比例进行独立、随机抽样，再将各层级的样本数据结合，得到总体有效样本。

除企业样本数据外，本次调查收集汇总数据资源相关公开数据，确保数据全面覆盖关键影响因素。利用机器学习模型，量化调查指标对数据资源发展水平的影响程度，为各项指标综合赋权，按照加权方式综合计算各关键指标。

总量类指标。将企业规模划分为大型企业、中型企业和微型企业，覆盖不同区域、不同行业及规模特征。对各层级独立抽样，结合各层级的企业总数（工商数据），加权推算总体规模。

比例类指标。采用分层比例估计法，以各层总体规模为权重进行加权汇总，推算总体比例。

增速类指标。为消除样本更替、口径变动带来的统计偏差，测算时首先筛选连续两个统计周期均进入有效样本的固定观测组，剔除期间新设、数据断档的不可比样本，再按分类分层维度得出全国总体增速。

全国数据资源统计调查工作组

电话: 010-88686123、88685702

传真: 010-68632953