

击破业务落地要害

中国面向人工智能的数据治理 行业研究报告

©2022.3 iResearch Inc.



前言-数据与数据治理：如今数据不再局限于传统数字形式的认知，由结构化数据延伸到半结构化、非结构化的数据范畴。**数据治理越来越受到企业的普遍重视**，在数据生命周期的各个阶段通过相应的工具与方法论，使数据发挥出更大的价值，**是实现数据服务与应用必不可少的阶段。**



主题-面向人工智能的数据治理：AI技术创新应用走向大规模落地，带动了大数据智能市场的蓬勃发展。**2021年大数据智能市场规模约为553亿元。**目前传统数据治理体系多停留在结构性数据化治理工作，尚难满足AI应用对数据的高质量要求。**企业可吸收传统体系的智慧沉淀，以AI应用数据需求为核心，优化建设“面向人工智能的数据治理”体系，显著提升AI应用的规模化落地效果。**



参与-行业规模与受益圈立足点：数据治理与AI应用产品开始交汇融合，厂商参与更加多元，咨询公司、数据服务提供商和人工智能产品服务商三方阵营构建行业竞争格局，而“智”，即AI应用，为面向人工智能的数据治理服务的核心立足点。**2021年面向人工智能的数据治理市场规模约为40亿元，预计五年后规模将突破百亿。**



实践-高频高价值应用及数据痛点：本篇报告选择**金融、零售、医疗和工业**四大典型行业为切入点，分析呈现各行业的信息化建设阶段与高频高价值的AI应用场景，并基于高频高价值AI应用引发的数据治理需求，对面向人工智能的数据治理体系搭建给到建设指导。



展望-治理陷阱与趋势洞察：1) 企业需避免落入“数据埋点大而全”的治理陷阱；2) 供需两侧需共同**保证数据治理体系建设后的运营流转**；3) 企业需建立符合管理现状及发展需求的**数据安全治理框架**，确保数据全周期的安全与合规；4) **联邦学习技术**可带来数据安全合规线内的共同富裕；5) 数据的**“自治与自我进化”**成为未来数据处理发展的必由之路，为企业打造“治理+AI”体系的良性循环。

前言：数据与数据治理

1

主题：面向人工智能的数据治理

2

参与：行业规模与受益圈立足点

3

实践：高频高价值应用及数据痛点

4

案例：标杆企业与新锐势力

5

展望：治理陷阱与趋势洞察

6

数据：范围界定

信息经济的“货币”，早已不限于数字形式

数据的价值被不断认可，“数据资产化”已经成为了企业发展的重要组成部分。长期以来，数据被理解为以数字形式存储的信息，而目前技术可以测量更多的事件和活动，人们可以收集、存储并分析这些不被视为传统数据的各类信息，如邮件、图片、音视频等。数据可根据其特性及治理方法差异划分为内部数据与外部数据，结构化数据、非结构化数据与半结构化数据，元数据与主数据等。

企业数据的主要类型

智慧 产生辨析判断、 发明创造能力 知识 经过人为 解读和经验 充实的信息 信息 经过收集和 整理的数据 数据 抽象的内容，独立 时无意义	分类标准	数据类型	定义以及特征	举例
	数据治理 常用数据 类型	元数据	是描述数据的数据（描述性标签），描述了数据（如数据元素、数据模型）、相关概念（如业务流程、应用系统、软件代码、技术架构）以及他们之间的联系	实体型组织、客户、人员基本配置
		主数据	描述企业核心实体的一组一致而统一的标识符和拓展属性，实体可包括现有或潜在客户、产品、服务、员工、供应商、提供商、层次结构和会计科目表等	数据标准、业务术语、指标定义
		实时数据	是在收集后立即传递的信息，所提供信息的及时性没有延迟	实时OLAP场景下的数据
	按照数据 格式分类	结构化数据	可以存储在传统的关系型数据库中，用二维表结构来表达实现的数据，可以用关系型数据库存储	Excel表格、SQL数据库里的数据
		非结构化数据	形式相对不固定，不方便用数据库二维逻辑表来表现的数据，通常存储在非关系型数据库中，数据量通常较大	文本、图片、HTML、各类报表和音频、视频
		半结构化数据	介于结构化与非结构化之间，半结构化数据可以通过灵活的键值调整获取相应信息，且数据的格式不固定	日志文件、XML文档、JSON文档、Email等
	按照数据 来源分类	企业内部数据	在企业内部经营中产生的数据，在企业的业务流程中产生或在业务管理规定中定义，受企业经营影响	国家、币种、汇率
		企业外部数据	企业通过公共领域合规获得的数据，其产生、修改不受公司影响	合同、项目、组织

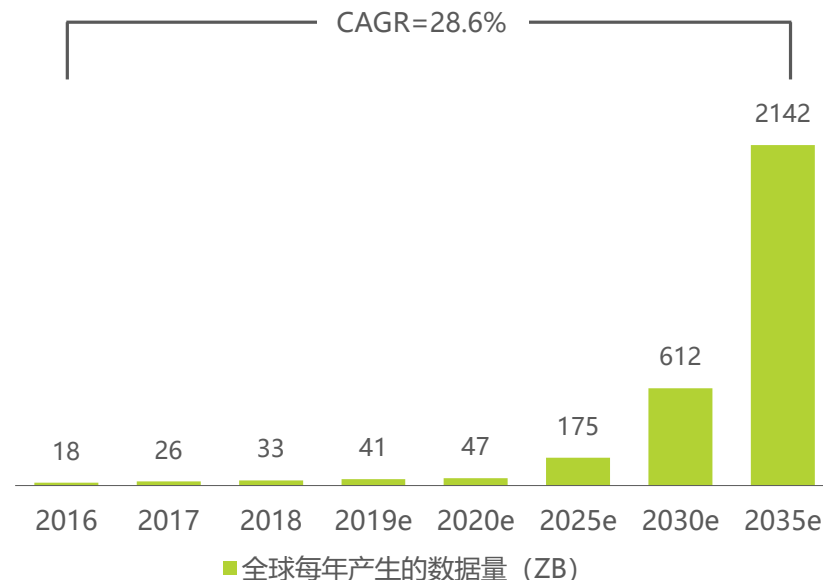
来源：艾瑞研究院自主研究绘制。

数据量：爆发式增长

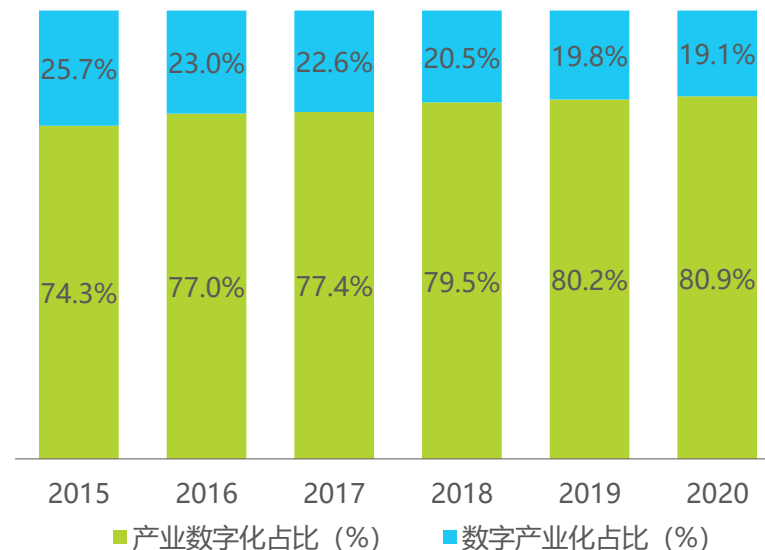
基础设施“扩容”、IoT 广泛连接带来的数据量暴涨

数据时代来临，数据量的暴涨为企业数字化提供了基础支撑，大量的业务数据能够被采集、存储并最终创造经济效益。数字化转型从头部企业的可选项，转变为更广泛企业的必选项。新变化为企业带来新机遇的同时，也带来了诸多挑战。很多企业在前期的信息化建设中，缺乏统筹规划，为解决当下业务问题而按照垂直的、个性化的业务逻辑独立采购与部署IT系统，导致企业内部形成多个数据孤岛。数据不规范、不一致、难以互联互通成为普遍问题，阻碍企业去充分发挥数据价值。这种先建设后治理的常态，使得数据治理越来越受到企业的普遍重视，另一方面，新兴技术与应用场景的快速落地，也带领数据治理需求在加速攀升。

2016-2035年全球产生的数据量



2015-2020年中国数字经济内部结构变化



注释：1ZB = 1024⁴ GB

来源：中国信通院，Statista (2020)，艾瑞研究院自主研究绘制。

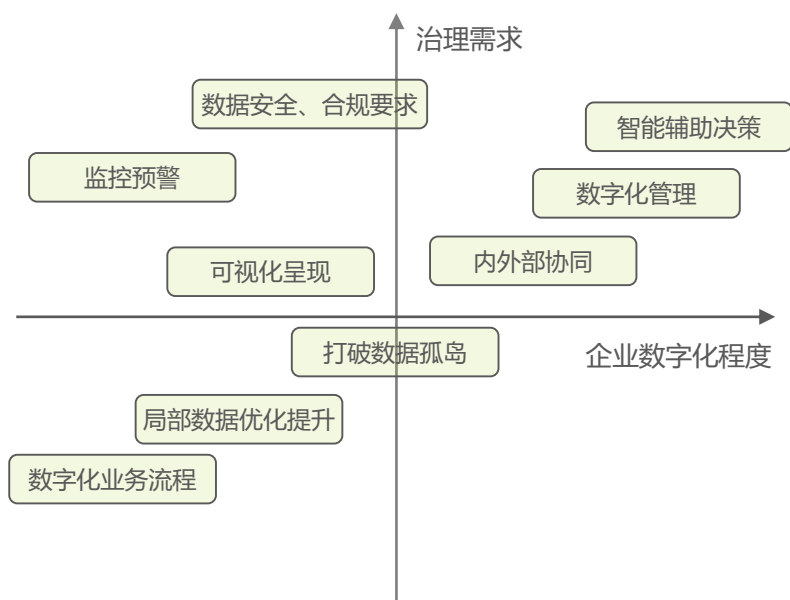
来源：中国信通院，艾瑞研究院根据专家访谈与公开资料研究绘制。

数据治理：需求释放

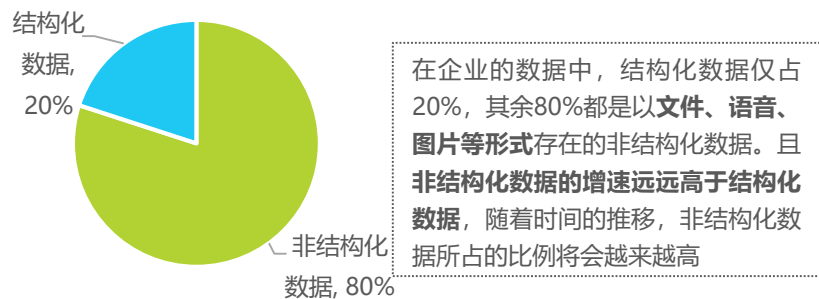
治理需求普遍存在，非结构化数据成为价值挖掘的重难点

企业历经数字化转型不同阶段时，需通过数据治理解决数据在生产、管理和使用中的问题，而数据治理的需求与复杂度也会随着企业数字化程度提升而增加。从企业内部的数据类型来看，非结构化数据占企业内数据总量的80%，却仅占整体使用率的30%，长期以来其价值未得到充分有效利用。未来，随着非结构化数据的积累增加与AI应用的数据需求推动，企业对非结构化数据的价值化需求将加速释放，而多源异构数据基础下的数据治理模块也将获得进一步的关注与优化。

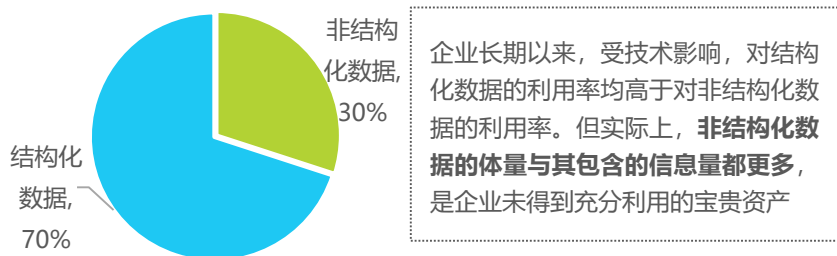
不同企业数字化程度下的主要数据治理需求



企业内结构化数据与非结构化数据占比情况



企业内结构化数据与非结构化使用现状



注释：仅列举代表性数据治理需求。
来源：艾瑞研究院自主研究绘制。

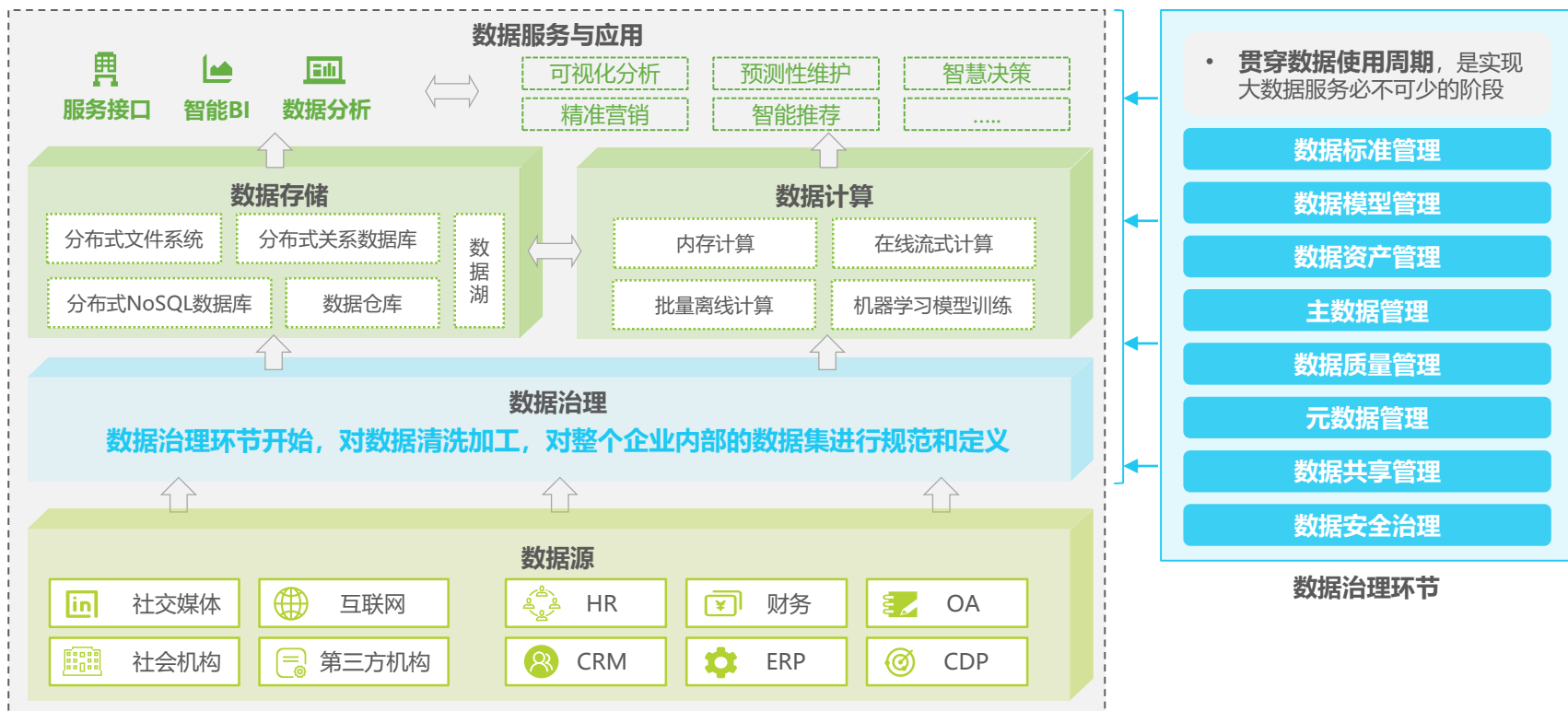
来源：艾瑞研究院根据专家访谈自主研究绘制。

数据治理：范围界定

数据治理为实现企业数据应用服务的重要环节

数据治理以数据源汇入为伊始，对数据进行清洗加工，并在数据存储、数据计算、数据服务应用等环节予以持续的治理服务，是企业实现数据服务与应用的重要环节。从数据层面来看，数据本身存在着从生产到消亡的生命周期，而数据治理会在数据生命周期的各阶段通过相应工具与方法论进行规范与定义，在企业内部构建出切实有效的数据闭环，使数据发挥出更大的价值。

数据治理在数据应用流程中的位置



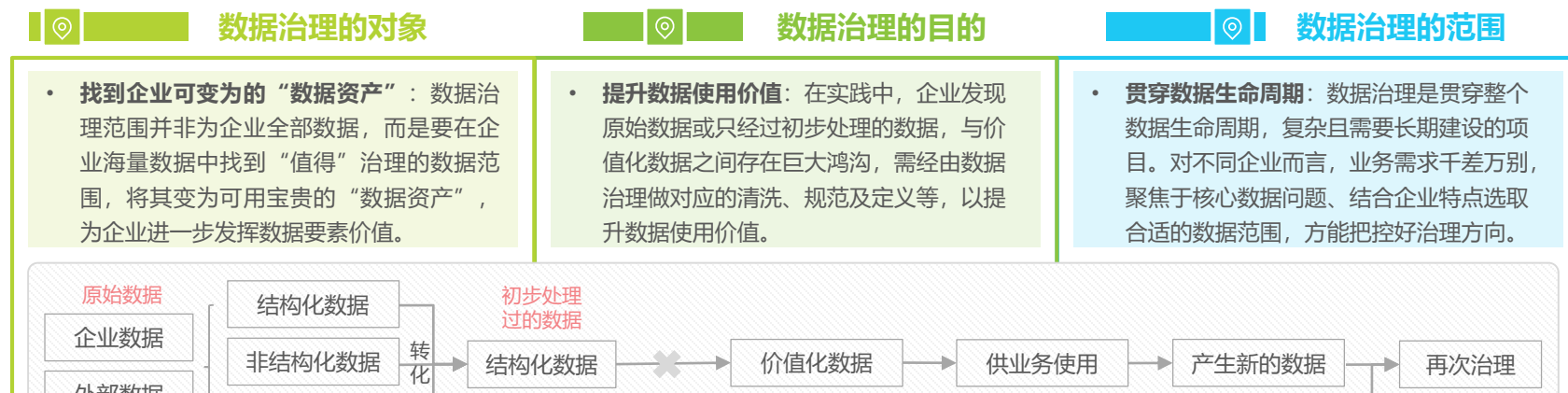
来源：艾瑞研究院自主研究绘制。

数据治理：整体概述

让数据可知、可用、可管，成为业务发展与创新的基石

数据治理旨在消除数据的不一致性，建立规范的数据标准，提高组织的数据质量与实现数据广泛共享，最终将数据变为宝贵资产，应用于企业的经营、管理与决策中。当下，让数据可知、可用、可管，充分发挥数据资产的价值已成为企业共同的数据治理目标。数据治理的对象与范围则会根据企业需求差异而有所区别。在不断发展变化的外部环境与业务需求下，企业数据治理工作在对应阶段也会有各自不同的目标。

数据治理的对象、目的与范围概述



预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=1_39834

