



中华人民共和国国家标准

GB/T XXXXX—XXXX

数据 基础术语

Data—Basic terminology

（征求意见稿）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX—XX—XX 发布

XXXX—XX—XX 实施

国家市场监督管理总局 发布
国家标准化管理委员会

目 次

前 言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

 3.1 基础通用 1

 3.2 数据基础设施 4

 3.3 数据资源 7

 3.4 数据技术 10

 3.5 数据流通 13

 3.6 融合应用 15

 3.7 安全保障 16

参 考 文 献 19

索 引 20

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由全国数据标准化技术委员会（SAC/TC 609）提出并归口。

本文件起草单位：全国科学技术名词审定委员会事务中心、中国电子技术标准化研究院、国家数据发展研究院、中国标准化研究院、国家发展和改革委员会创新驱动发展中心、国家信息技术安全研究中心、中国电子信息产业发展研究院、下一代互联网关键技术和评测北京市工程研究中心有限公司、上海计算机软件技术开发中心、北京大学、清华大学、中国人民大学、北京交通大学、华东师范大学、中国科学院信息工程研究所、贵阳大数据交易所有限责任公司、杭州市数据集团有限公司、中电数据产业集团有限公司、中电科大数据研究院有限公司、贵州大数据产业集团有限公司、内蒙古自治区大数据中心、中国移动通信集团有限公司、中国联合网络通信集团有限公司、中国石油化工集团有限公司、阿里巴巴（中国）有限公司、华为技术有限公司、蚂蚁科技集团股份有限公司、杭州数据交易所有限公司、陕西丝路数据交易所有限公司、湖南大数据交易所有限公司、广西北部湾大数据交易中心有限公司、奇安信科技集团股份有限公司、广东卓建律师事务所、浪潮电子信息产业股份有限公司、杭州安恒信息技术股份有限公司。

本文件主要起草人：梅宏、张晖、张群、刘金婷、余前帆、王晨、刘文强、郭建民、王亚沙、张向宏、王为中、邓成龙、陈蔚、蔡戟、罗松、王海涛、安小米、杜小勇、周傲英、李凤华、牛犇、班帅帅、张树礼、吴志刚、张宇、张天、杨韬、张旭东、宋俊典、刘格扬、赵俊峰、王天瑞、国丽、刘珏、范举、黄科满、肖连春、潘凯伟、彭晋、昌文婷、韩坤洁、顾延甲、王华、刘军、郭宽厚、吴坤、李金夏、王远、蒋楠、孙勇、马万钟、李翩翩、张梦雪、曹出、周海扬、李兰兰、楚赟、胡德鹏、梁小涛、丁熙。

数据 基础术语

1 范围

本文件界定了数据领域中的常用术语和定义。
本文件根据国家数据标准体系确定术语分类。
本文件适用于数据领域的教学、科研、生产、经营和技术交流。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

3.1 基础通用

3.1.1

数据 data

任何以电子或以其他方式对信息的记录。

3.1.2

数据科学 data science

运用数学、计算机等学科的研究方法，通过挖掘、分析及建模，研究数据规律，从数据中提取知识并支持决策的一门交叉学科。

3.1.3

数据价值 data value

数据在特定应用场景中通过分析、处理或流通所体现的经济利益、社会效益或科学价值。其核心在于通过数据资源的开发利用，实现支撑决策、优化流程、创新服务等直接或间接效益。

3.1.4

数据资源 data resources

具有价值创造潜力的数据的总称，通常指以电子化形式记录和保存、可机器读取、可供社会化再利用的数据集合。

3.1.5

数据要素 data factor

投入到生产经营活动、参与价值创造的数据资源。

3.1.6

数据资产 data assets

特定主体合法拥有或者控制的，能进行货币计量的，且能带来经济利益或社会效益的数据资源。

[来源：GB/T 46353—2025，3.1，有修改]

3.1.7

数据治理 data governance

提升数据的质量、安全、合规性，推动数据有效利用的过程。

[来源：GB/T 36073—2025, 3.2, 有修改]

3.1.8

数据管理 data management

数据资源获取、控制、价值提升等活动的集合。

[来源：GB/T 34960.5—2018, 3.2]

3.1.9

数据标准 data standard

数据的命名、定义、结构和取值规范方面的规则和基准。

[来源：GB/T 36344—2018, 2.8]

3.1.10

数据开放 data opening

数据可以被合法用户使用、再利用、再分发的机制。

3.1.11

数据共享 data sharing

在一定的条件下，数据拥有方将数据与相关方共同使用。

注：根据不同共享类型，分为无条件共享、有条件共享以及不予共享。

[来源：GB/T 36625.4—2021, 2.2]

3.1.12

数据流通 data circulation

数据在不同主体之间流动的过程，包括数据开放、共享、交易、交换等。

3.1.13

数据利用 data use

为特定目的对数据进行处理和管理，使数据价值得以实现的行为过程、方法、手段和服务的一组能力。

[来源：GB/T 42458—2023, 3.3, 有修改]

3.1.14

数据产品 data product

自然人、法人或者非法人组织对其合法获取的数据资源，经过实质性加工和创新性劳动后形成的，可满足特定需求的数据加工品和数据服务。

3.1.15

数据服务 data service

基于数据资源开发的，旨在优化数据利用的数据解决方案。

3.1.16

数据产业 data industry

利用现代信息技术对数据资源进行产品或服务开发，并推动其流通应用所形成的产业，包括数据采集汇聚、计算存储、流通交易、开发利用、安全治理和数据基础设施建设等。

3.1.17

数字经济 digital economy

以数据资源作为关键生产要素、以现代信息网络作为重要载体、以信息通信技术的有效使用作为效率提升和经济结构优化的重要推动力的一系列经济活动。

3.1.18

数字产业化 digital industrialization

移动通信、人工智能等数字技术向数字产品、数字服务转化，数据向资源、要素转化，形成数字新产业、新业态、新模式的过程。

3.1.19

产业数字化 industrial digitization

传统的农业、工业、服务业等产业通过应用数字技术、采集融合数据、挖掘数据资源价值，提升业务运行效率，降低生产经营成本，进而重构思维认知，整体性重塑组织管理模式，系统性变革生产运营流程，不断提升全要素生产率的过程。

3.1.20

数据安全 data security

通过采取必要措施，确保数据处于有效保护和合法利用的状态，以及具备保障持续安全状态的能力。

[来源：GB/T 45577—2025，3.2]

3.1.21

数据基础设施 data infrastructure

从数据要素价值释放的角度出发，面向社会提供数据采集、汇聚、传输、加工、流通、利用、运营、安全服务的一类新型基础设施，是集成硬件、软件、模型算法、标准规范、机制设计等在内的有机整体。

3.1.22

数据生命周期 data lifecycle

数据从采集到销毁的整个过程，包括数据的采集、存储、传输、使用、销毁等阶段。

3.1.23

数字连续性 digital continuity

确保数据在生命周期内具有真实性、可靠性、完整性、可用性的数据资源持续性管理过程。

3.1.24

数据质量 data quality

在指定条件下使用时，数据的特性满足明确的和隐含的要求的程度。主要包括数据的完整性、真实性、可靠性、及时性、一致性、可追溯性和包容性。

[来源：GB/T 36344—2018，2.3，有修改]

3.1.25

完整性 completeness

数据完备的特性，数据项以及数据值存在和缺失的程度。

[来源：GB/T 42458—2023，6.2表2，有修改]

3.1.26

真实性 authenticity

数据实体与事实的符合程度。

[来源：GB/T 42458—2023，6.2表2，有修改]

3.1.27

可靠性 reliability

数据对于预期用途具有支撑证明作用的程度。

[来源：GB/T 42458—2023，6.2表2，有修改]

3.1.28

及时性 **timeliness**

数据提供符合利益相关方需求的时限程度。

[来源：GB/T 42458—2023，6.2表2，有修改]

3.1.29

一致性 **consistency**

不同系统和设备存储或使用的相同数据一致的程度，以及不同数据服务中存储的相同数据同时被修改的程度。

[来源：GB/T 42458—2023，6.2表2，有修改]

3.1.30

可追溯性 **traceability**

对源头数据演变为当前数据状态的过程有完整记录并实现其历史、使用情况和所处位置可查的程度。

[来源：GB/T 42458—2023，6.2表2，有修改]

3.1.31

包容性 **inclusiveness**

对信息弱势群体公共数据权益保障的程度。

[来源：GB/T 42458—2023，6.1表1，有修改]

3.1.32

词元 **token**

人工智能领域智能设备中信息存储、处理和交换的具有一定语义的基本符号表达。在大模型中作为模型处理和交换信息的最小单位，具有可计量、可定价、可交易的特征。

3.2 数据基础设施

3.2.1

数据中心 **data center**

由计算机场站（机房）、机房基础设施、信息系统硬件（物理和虚拟资源）、信息系统软件、信息资源（数据）和人员以及相应的规章制度组成的组织。

[来源：GB/T 33136—2016，3.1.1]

3.2.2

数据存储 **data storage**

将数据用某种结构保存在计算机的物理介质中，以便加工、重用或存档。

3.2.3

云存储 **cloud storage**

整合多种网络存储资源，以存储服务的形式提供给用户使用的一种存储模式。

3.2.4

数据仓库 **data warehouse**

一种可扩展的数据存储架构，通常采用结构化数据模型，用于存储和管理经过清洗、集成和转换的数据。

3.2.5

数据湖 data lake

一种高度可扩展的数据存储架构，它专门用于存储大量原始数据和衍生数据，这些数据可以来自各种来源并以不同的格式存在，包括结构化、半结构化和非结构化数据。

3.2.6

湖仓一体 data lakehouse

一种新型的开放式的存储架构，打通了数据仓库和数据湖，将数据仓库的高性能及管理能力和数据湖的灵活性融合起来，底层支持多种数据类型并存，能实现数据间的相互共享，上层可以通过统一封装的接口进行访问，可同时支持实时查询和分析。

3.2.7

可信数据空间 trustworthy data space

基于共识规则，联接多方主体，实现数据资源共享共用的一种数据流通利用基础设施，是数据要素价值共创的应用生态，是支撑构建全国一体化数据市场的重要载体。可信数据空间须具备数据可信管控、资源交互、价值共创三类核心能力。

3.2.8

数据计算 data computing

对数据进行算术、逻辑等运算以分析或提取信息的过程。

3.2.9

云计算 cloud computing

由位于网络中央的一组服务器把其计算、存储、数据等资源以服务的形式提供给请求者，以完成信息处理任务的方法和过程。

注：在此过程中，被服务者只是提供需求并获取服务结果，对于需求被服务的过程并不知情。

3.2.10

分布式计算 distributed computing

利用分散在不同位置的多台计算机通过消息传递（计算机间通信）协同工作以完成计算任务的方法和过程。

3.2.11

分布式文件系统 distributed filesystem

将数据资源存储在分布式网络结点上，提供统一的文件访问接口的一种文件系统。

3.2.12

算力 computing power; computility

图形处理器（GPU）、中央处理器（CPU）等设备执行计算密集型任务的计算能力。

3.2.13

算力规模 computing power scale; computility scale

组织为执行计算密集型任务而自行构建的总算力的大小。

3.2.14

算力池化 computing power pooling; computility pooling

通过算力虚拟化和应用容器化等关键技术，对各类异构、异地的算力资源与设备进行统一注册和管理，实现对大规模集群内计算资源的按需申请与使用。

3.2.15

算力调度 computing power scheduling; computility scheduling

基于用户业务需求匹配算力资源，将业务、数据、应用调配至匹配的算力资源池进行计算，实现计算资源合理利用。

3.2.16

“东数西算”工程 East Data West Computing Project

把东部地区经济活动产生的数据和需求放到西部地区计算和处理，对数据中心在布局、网络、电力、能耗、算力、数据等方面进行统筹规划的重大工程。

3.2.17

全国一体化算力网 national integrated computing power network; national integrated computility network

以信息网络技术为载体，促进全国范围内各类算力资源高比例、大规模一体化调度运营的数字基础设施。作为“东数西算”工程的2.0版本，具有集约化、一体化、协同化、价值化四个典型特征。

3.2.18

数联网 internet of data

由数据流通接入终端、数据流通网络、数据流通服务平台构成，提供一点接入、广泛连接、标准交付、安全可信、合规监管、开放兼容的数据流通服务。

3.2.19

高速数据网 high speed data network

面向数据流通利用场景，依托网络虚拟化、软件定义网络（SDN）等技术，提供弹性带宽、安全可靠、传输高效的数据传输服务。

3.2.20

物联网 internet of things; IoT

基于感知控制设备，通过通信网络，使物理实体、人、系统和信息资源相连接，响应和处理物理和虚拟世界信息的基础设施。

[来源：GB/T 33745—2025，3.1.1]

3.2.21

物联感知数据 IoT sensing data

通过物联网终端设备（如传感器、射频识别设备、摄像头等）对物理世界实时监测所采集的原始数据或初步处理数据，通常包含环境参数、设备状态、位置信息等动态信息。

注：典型特征为实时性、海量性、异构性，需依赖边缘计算或云计算平台进行进一步整合与分析。

3.2.22

产业互联网 industrial internet

利用数字技术、数据要素推动全产业链数据融通，赋能产业数字化、网络化、智能化发展，推动业务流程、组织架构、生产方式等重组变革，实现产业链上下游协同转型、线上线下融合发展、全产业降本增效与高质量发展，进而形成新的产业协作、资源配置和价值创造体系。

3.2.23

数场 data field

依托开放性网络、算力和隐私保护计算、区块链等各类关联功能设施，面向数据要素提供线上线下资源登记、供需匹配、交易流通、开发利用、存证溯源等功能，支持多场景应用的一种综合性数据流通利用设施。

3.2.24

数据元件 data component

连接数据供需两端的中间态，将原始数据与数据应用解耦，提供统一标准、自主可控、安全可靠、全程监管的数据存储和加工服务，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。

3.2.25

区块链 blockchain

分布式网络、加密技术、智能合约等多种技术集成的新型数据库软件，具有多中心化、共识可信、不可篡改、可追溯等特性，主要用于解决数据流通过程中的信任和安全问题。

3.3 数据资源

3.3.1

数据分类 data classification

按照一定的标准和方法对数据进行归类和划分类别的过程。

3.3.2

数据分级 data grading

结合数据分类，综合考虑数据的重要性、敏感性、泄露风险及可能对国家安全、公共利益或个人、组织合法权益造成的危害程度等因素，将数据划分成不同安全级别的过程，其核心是为不同级别的数据采取相适应的保护措施，确保数据安全。

3.3.3

公共数据 public data

各级党政机关、企事业单位依法履职或提供公共服务过程中产生的数据。

3.3.4

公共数据资源 public data resources

各级党政机关、企事业单位依法履职或提供公共服务过程中产生的具有利用价值的数据集合。

3.3.5

政务数据 government data

各级政务部门及其技术支撑单位在履行职责过程中依法采集、生成、存储、管理的各类数据资源。

注：根据可传播范围，政务数据一般包括可共享政务数据、可开放公共数据及不宜开放共享政务数据。

[来源：GB/T 38664.1—2020，3.1]

3.3.6

科学数据 scientific data

在科学研究、实验观测、模拟计算等活动中产生的，用于揭示自然规律或验证科学假设的原始数据、衍生数据及配套元数据等。

3.3.7

企业数据 enterprise data

企业在生产经营过程中形成或合法获取、持有的数据。

3.3.8

个人信息 personal information

以电子或者其他方式记录的与已识别或者可识别的自然人有关的各种信息。

[来源：GB/T 45574—2025, 3.1]

3.3.9

个人信息数据 personal information data

个人数据

承载个人信息的数据。

3.3.10

敏感个人信息 sensitive personal information

一旦泄露或非法使用,容易导致自然人的人格尊严受到侵害或者人身财产安全受到危害的个人信息。

注：敏感个人信息包括生物识别、宗教信仰、特定身份、医疗健康、金融账号、行踪轨迹等信息和不满十四周岁未成年人的个人信息。

[来源：GB/T 45574—2025, 3.2]

3.3.11

结构化数据 structured data

一种数据的表示形式,按此种形式,由数据元素汇集而成的每个记录的结构都是一致的并且可以使用关系模型予以有效描述。这些数据具有清晰的字段和属性,每个字段都有特定的数据类型和取值范围。

3.3.12

半结构化数据 semi-structured data

不符合关系型数据库或其他数据表的形式关联起来的数据模型结构,但包含相关标记,用来分隔语义元素以及对记录和字段进行分层的一种数据化结构形式。

3.3.13

非结构化数据 unstructured data

不具有预定义模型或未以预定义方式组织的数据。

3.3.14

多模态数据 multi-modal data

包含结构化数据、半结构化数据或非结构化数据等多种形态的数据。

[来源：GB/T 42135—2022, 3.1, 有修改]

3.3.15

大数据 big data

具有体量巨大、来源多样、生成极快、多变等特征并且难以用传统数据体系结构有效处理的包含大量数据集的数据。

注：国际上,大数据的4个特征普遍不加修饰地直接用volume、variety、velocity和variability予以表述,并分别赋予了它们在大数据语境下的定义：

- 1) 体量 volume：构成大数据的数据集的规模。
- 2) 多样性 variety：数据可能来自多个数据仓库、数据领域或多种数据类型。
- 3) 速度 velocity：单位时间的数据流量。
- 4) 多变性 variability：大数据其他特征,即体量、速度和多样性等特征都处于多变状态。

[来源：GB/T 35295—2017, 2.1.1]

3.3.16

主数据 master data

组织中需要跨系统或跨部门进行共享的核心业务实体数据。

[来源：GB/T 36073—2025, 3.16]

3.3.17

元数据 metadata

定义和描述特定数据的数据，提供了关于数据的结构、特征和关系的信息，有助于组织、查找、理解、管理数据。

3.3.18

原始数据 raw data

初次产生或源头采集的、未经加工处理的数据。

3.3.19

衍生数据 derivative data; derived data

数据处理者对其享有使用权的数据，在保护各方合法权益前提下，通过利用专业知识加工、建模分析、关键信息提取等方式实现数据内容、形式、结构等实质改变，从而显著提升数据价值，形成的数据。

3.3.20

动态数据 data in motion

处于活动状态，典型特征表现为数据的速度和多变性特征的数据。

注：动态数据在网络上传输或暂时驻留于计算机内存中供读取或更新，以实时或近实时方式进行处理和分析。

[来源：GB/T 35295—2017, 2.1.36]

3.3.21

静态数据 data at rest

处于静止状态，典型特征表现为数据的体量和多样性特征的数据。

[来源：GB/T 35295—2017, 2.1.37]

3.3.22

热数据 hot data

需要被实时或近实时访问以支撑核心业务运行、直接影响关键决策或用户体验的数据。通常需要高性能存储保障，管理成本最高。

3.3.23

温数据 warm data

访问频率和时效性要求适中，用于支持周期性分析、报表生成或非实时业务的数据。通常采用性能与成本均衡的存储策略。

3.3.24

冷数据 cold data

主要用于合规性归档、历史追溯或低频度离线分析的数据。对访问延迟不敏感，通常采用低成本存储方案以优化总体拥有成本。

3.3.25

数据目录 data catalogue

用于描述数据特征的一组信息，为用户提供对数据的全面了解和访问途径，以提高数据的可发现性、可理解性和可管理性。本文件中的数据目录包括数据资源目录和数据产品目录。

3.3.26

数据集 data set

具有一定主题，可以标识并可以被计算机化处理的数据集合。

[来源：GB/T 36344—2018，2.6]

3.3.27

高质量数据集 high-quality dataset

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。

3.3.28

开放数据集 open dataset

由政府、企业或机构主动公开，供社会公众自由访问、使用和分发的结构化或非结构化数据集，通常以标准格式并通过开放许可协议发布。

3.4 数据技术

3.4.1

数据处理 data handling

数据操作的系统执行，以实现特定目的的数据收集、存储、使用、加工、传输、提供、公开等活动。

[来源：GB/T 35274—2023，3.2，有修改]

3.4.2

数据收集 data collection

通过向数据源方采集和捕获信息，以创建数据集或获取关于特定现象的知识的過程。

3.4.3

数据汇聚 data convergence

将来自多个异构、分散数据源的数据，通过物理或逻辑方式集中到统一数据平台或存储中的过程。

3.4.4

数据清洗 data cleaning

通过识别错误、冗余或不一致数据并修复或移除的过程。

3.4.5

数据加工 data processing

通过数据变换、数据转换、数据编码、数据计算、数据压缩、数据分析等数据操作，生成新数据（集）的数据处理活动。

[来源：GB/T 35274—2023，3.6]

3.4.6

数据分析 data analysis

通过特定的技术和方法，对数据进行整理、研究、推理和概括总结，从数据中提取有用信息、发现规律、形成结论的过程。

3.4.7

数据挖掘 data mining

数据分析的一种手段，是通过统计分析、机器学习、模式识别、专家系统等技术，挖掘出隐藏在数据中的信息或者价值的过程。

3.4.8

数据融合 data fusion

基于一组或多组数据，通过一定的处理过程以获得新的或更高质量信息的过程。

[来源：GB/T 33745—2017，2.5.6]

3.4.9**数据特征 data characteristic**

能准确描述数据本质的核心信息元素。

3.4.10**数据特征层次 data characteristic hierarchy**

从不同粗细粒度表征数据特征的数据层次结构。

[来源：GB/T 35295—2017，2.1.51]

3.4.11**数据可视化 data visualization**

通过统计图表、图形、地图等图形化手段，将数据中包含的有用信息清晰有效地传达出来，以便于数据使用者更好地理解和分析数据。

3.4.12**数据标识 data identifier**

一种资源互通技术，通过为数据资源分配唯一标识符，实现快速准确的数据检索和定位，实现数据全生命周期的可追溯性和可访问性。

3.4.13**数据标注 data labeling; data annotation**

给数据样本指定目标变量和赋值的过程。

[来源：GB/T 42755—2023，3.1，有修改]

3.4.14**数据标签 data label**

用于标识数据特性或分类的数据项，帮助进行数据分类、检索和共享。

3.4.15**标识符 identifier**

微数据中的一个或多个属性，可以实现对个人信息主体的唯一识别。

[来源：GB/T 42460—2023，3.5]

3.4.16**直接标识符 direct identifier**

微数据中的属性，在特定环境下可以单独识别个人信息主体。

[来源：GB/T 42460—2023，3.6]

3.4.17**准标识符 quasi-identifier**

微数据中的属性，结合其他属性可唯一识别个人信息主体。

[来源：GB/T 42460—2023，3.7]

3.4.18**数据接入 data access**

通过技术手段将分散的数据源与目标系统或平台建立连接，实现数据的采集、传输与初步整合的过程。

3.4.19

接入连接器 access connector

连接接入主体与接入主体、接入主体与业务节点、接入主体与区域/行业功能节点的规范化软硬件系统，数据供需双方均可通过接入连接器接入数据基础设施。

3.4.20

数据传输 data transmission

数据在不同实体之间进行交换或流动的过程，包括本地网络传输、远程通信等。

3.4.21

数据交换 data exchange

通过采用约定的信息格式、控制协议和开放接口，在不同主体之间相互传递数据，以实现不同系统间通信、信息共享、协同运作的过程。

3.4.22

数据交换模型 data exchange model

定义不同系统间结构化数据传输规则与流程的标准化框架。

3.4.23

数据交互 data interaction

主体之间以数据的形式进行交流和协作的过程。

3.4.24

数据应用 data application

基于数据构建的可直接服务于用户的软件或服务。

3.4.25

数据运营 data operation

对数据进行采集、整理、分析、挖掘和利用等系统性管理，通过制定策略、流程和技术手段，从数据中提取有价值的信息，以实现优化业务流程、提升运营效率、支持决策制定、创造商业价值和社会效益的一系列活动。

3.4.26

数据销毁 data destruction

从存储介质或系统中永久且安全地擦除数据，使其无法恢复，防止未经授权访问的活动。

3.4.27

机器学习 machine learning

通过计算技术优化模型参数的过程，使模型的行为反映数据或经验。

[来源：GB/T 41867—2022，3.2.10]

3.4.28

知识图谱 knowledge graph

以结构化形式描述的知识元素及其联系的集合。

[来源：GB/T 42131—2022，3.6]

3.4.29

大模型 large model

基于大量数据训练得到，具有复杂计算架构，能处理复杂任务，且具备一定泛化性的深度学习模型。

注：大模型的参数量由其功能和模态决定，一般不低于1亿。大模型训练使用的数据总量受参数量的影响，达到收敛的大模型的参数量的对数与其训练数据总量的对数成正比。

[来源：GB/T 45288.1—2025，3.1]

3.5 数据流通

3.5.1

数据流通环境 data circulating environment

支持数据在不同主体之间进行传输、存储、交换、开放、共享、交易、计算等流通利用活动的硬件、软件设施和服务的集合。

3.5.2

数据合规 data compliance

在数据全生命周期中，确保数据资源、数据产品和服务的来源、内容、可交易性等符合法律法规的相关规定。

3.5.3

数据产权 data property rights

权利人对特定数据享有的财产性权利，包括数据持有权、数据使用权、数据经营权等。

3.5.4

数据持有权 data holding rights

权利人自行持有或委托他人代为持有合法获取的数据的权利。旨在防范他人非法违规窃取、篡改、泄露或者破坏持有人持有的数据。

3.5.5

数据使用权 data use rights

权利人通过加工、聚合、分析等方式，将数据用于优化生产经营、形成衍生数据等的权利。一般来说，使用权是权利人在不对外提供数据的前提下，将数据用于内部使用的权利。

3.5.6

数据经营权 data operation rights

权利人通过转让、许可、出资或者依法设立担保等有偿或无偿的方式对外提供数据的权利。

3.5.7

数据产权登记 data property rights registration

数据产权登记机构按照统一的规则对数据的来源、描述、内容等进行审查，记载数据权利归属等信息，并出具登记凭证的行为。

3.5.8

数据产权登记机构 data property rights registration authority

由各级数据主管部门设立或指定的，提供数据相关登记服务的机构。

3.5.9

数据产权登记平台 data property rights registration system

支撑数据产权登记全流程服务管理的信息化系统。

3.5.10

公共数据资源登记主体 public data resources registration entity

根据工作职责直接持有或管理公共数据资源的单位，以及依法依规对授权范围的公共数据资源进行开发运营的法人组织。

3.5.11

公共数据资源授权运营 authorized operation of public data resources

将县级以上地方各级人民政府、国家行业主管部门持有的公共数据资源，按照法律法规和相关要求，授权符合条件的运营机构进行治理、开发，并面向市场公平提供数据产品和技术服务的活动。

3.5.12

数据交易 data trading

数据供方和需方之间进行的，以特定形态数据为标的，以货币或者其他等价物作为对价的交易行为。

3.5.13

数据流通服务机构 data circulation service agency

依法成立，连接数据供需双方，促进数据流通交易，推进数据要素市场化价值化的法人组织。

3.5.14

数据交易所 data exchange center

数据交易中心

经政府部门批准或主导成立，为数据流通交易活动提供场所和基础设施，组织开展数据流通交易的法人组织。

3.5.15

数据流通服务平台企业 data circulation service platform enterprise

依法成立，依托产业互联网平台、数据基础设施、云服务平台等载体，开展数据流通交易的法人组织。

3.5.16

数据商 data vendor

汇聚数据资源或数据产品，开展数据产品和服务的开发、销售等业务的法人组织。

3.5.17

第三方专业服务机构 third-party professional service organization

为促进数据交易活动合规高效开展，提供数据集成、质量评价、质量认证、数据经纪、合规认证、安全审计、数据公证、数据保险、数据托管、资产评估、争议调解、风险评估、人才培养、咨询服务等第三方服务的专业化组织。

3.5.18

数据供方 data supplier

数据供给方

数据提供方

产生、持有或控制数据，并在流通中出售、提供数据的组织或个人。

注：数据加工处理流通情形可能涉及多个提供方对数据进行联合处理或融合计算，针对非本方提供的数据而言，提供方也属于数据接收方。

3.5.19

数据需方 data demander

数据需求方

在数据流通中采集、接收、购买或使用数据的组织或个人。

3.5.20

数据交易额 data transaction volume

数据产品和服务的交易总金额。不包括算力、信息化系统采购、智慧城市项目采购等交易的金额。

3.5.21

交易笔数 transaction count

统计周期内，买卖双方开展数据交易活动，签订交易合同（含续约）的数量。

3.5.22

数据交易撮合 data trading matching

帮助数据供方、需方达成数据交易的行为。

3.5.23

数据交付 data delivery

数据供需双方达成协议或形成交易一致意见后，数据供方将标的以一定的形式提供给数据需方的过程。

3.5.24

数据交付平台 data delivery platform

为数据交易双方或多方提供数据交易撮合及相关服务的信息系统，提供可信数据交易服务及运营支撑，根据供需方数据交易合约，进行数据调度，实现数据产品及服务交付全过程安全可信。

3.5.25

数据跨境流入 cross-border data inflow

境外数据通过网络或物理载体传输或带入至本国境内存储、处理或使用的行为。

3.5.26

数据跨境流出 cross-border data outflow

本国数据向境外传输、存储或提供访问的行为。

3.6 融合应用

3.6.1

数字产业集群 digital industry cluster

以数据要素驱动、数字技术赋能、数字平台支撑、产业融通发展、集群生态共建为主要特征的产业组织新形态。

3.6.2

数据标注产业 data annotation industry

对数据进行筛选、清洗、分类、注释、标记和质检等加工处理的产业。

3.6.3

数据标注人员 data labeler

数据标注员

承担数据标注工作的专业人员。

3.6.4

数据要素市场化配置 market-oriented allocation of data factor

通过市场机制来配置数据要素，旨在建立一个更加开放、安全、合规和高效的数据流通环境，不断释放数据要素价值。

3.6.5

应用场景 application scenario

数据产品在特定领域中的实际使用环境和背景，能直接影响数据产品功能定义、技术要求、质量标准和数据价值。

3.6.6

城市全域数字化转型 citywide digital transformation

城市以全面深化数据融通和开发利用为主线，综合利用数字技术和制度创新工具，实现技术架构重塑、城市管理流程变革和产城深度融合，促进数字化转型全领域增效、支撑能力全方位增强、转型生态全过程优化的城市高质量发展新模式。

3.6.7

智慧城市 smart city

运用信息通信技术，有效整合各类城市管理系统，实现城市各系统间信息资源共享和业务协同，推动城市管理和智慧化，提升城市运行管理和公共服务水平，提高城市居民幸福感和满意度，实现可持续发展的一种创新型城市。

注1：ISO（国际标准化组织）将智慧城市定义为：“在已建环境中对物理系统、数字系统和人类系统进行有效整合，从而为市民提供一个可持续的、繁荣的包容性的未来”。

[来源：ISO/IEC 30182:2017, 2.14]

注2：ITU-T（国际电信联盟电信标准化部门）强调可持续发展，将智慧可持续发展城市（smart sustainable city）定义为：“使用信息通信技术和其他手段来改善生活质量、提高城市运营和服务效率以及城市竞争力，同时确保满足当代和后代的经济、社会、环境和文化方面需求的一种创新型城市”。

[来源：ITU-T Y.4900/L.1600, 3.2.2]

3.6.8

数字孪生 digital twin

具有保证物理状态和虚拟状态之间以适当速率和精度同步的数据连接的特定目标实体的数字化表达。

[来源：GB/T 43441.1—2023, 3.4]

3.6.9

数字消费 digital consumption

数字技术、应用支撑形成的消费活动和消费方式，既包括对数智化技术、产品和服务的消费，也包括消费内容、消费渠道、消费环境的数字化与智能化，还包括线上线下深度融合的消费新模式。

3.7 安全保障

3.7.1

数据安全技术 data security technology

在数据收集、存储、使用、加工、传输、提供、公开等处理活动中所采用的技术手段和方法。

3.7.2

数据安全能力 data security capability

通过技术与管理措施保障数据处理活动的有效保护、合法利用及持续安全的能力。

3.7.3**可信执行环境 trustworthy execution environment**

基于硬件级隔离及安全启动机制，为确保安全敏感应用相关数据和代码的机密性、完整性、真实性和不可否认性目标构建的一种软件运行环境。

[来源：GB/T 41388—2022，3.3]

3.7.4**可信身份凭证 trustworthy identity credential**

一种基于数字加密等技术的身份凭证，它允许个人或组织以安全、隐私保护的方式，证明和验证其身份、资格、属性或其他信息。

3.7.5**失效身份凭证 invalid credential**

凭据无效或已过期的身份凭证，包括资格、属性或其他信息（扩展）变更导致的身份失效。

3.7.6**数据脱敏 data desensitization**

通过一系列数据处理方法对原始数据进行处理以屏蔽敏感数据的一种数据保护方法。

注：数据脱敏侧重于屏蔽敏感字段，而去标识化侧重于消除个人信息的可识别性，二者均属于隐私保护技术，可结合使用。

[来源：GB/T 37988—2019，3.12；GB/T 35273—2020，3.15]

3.7.7**去标识化 de-identification**

个人信息经过处理，使其在不借助额外信息的情况下无法识别特定自然人的过程。

注：去标识化建立在个体基础之上，保留了个体颗粒度，采用假名、加密、哈希函数等技术手段替代对个人信息的标识。

[来源：GB/T 35273—2020，3.15]

3.7.8**匿名化 anonymization**

个人信息经过处理无法识别特定自然人且不能复原的过程。

注1：匿名化处理后的信息也称匿名数据，其不属于个人信息。

注2：匿名化是去标识化的极端情况，强调无法识别特定自然人且不能复原。

[来源：GB/T 35273—2020，3.14，有修改]

3.7.9**重标识 reidentification**

把去标识化的数据集重新关联到原始个人信息主体或一组个人信息主体的过程。

[来源：GB/T 42460—2023，3.8]

3.7.10**一致性验证 consistency verification**

验证交易数据与原始数据在格式、内容及完整性上的一致性，确保数据在传输、存储过程中未被篡改或损坏的技术手段。

[来源：GB/T 25069—2022，3.7.16，有修改]

3.7.11

用户授权令牌 user authorization token

一个安全且不可预测的加密字符串或数字序列，由系统生成并分配给特定的用户或应用程序，作为在有效期内有效的临时凭证，用于验证用户身份并授权其访问受保护的资源或服务。

3.7.12

安全多方计算 secure multi-party computation; SMPC

在一个分布式网络中，多个参与实体各自持有秘密数据，各方希望以这些数据为输入共同完成对某函数的计算，而要求每个参与实体除计算结果、预期可公开的信息外均不能得到其他参与实体的任何输入信息。主要研究针对无可信第三方情况下，安全地进行多方协同的计算问题。

3.7.13

联邦学习 federated learning

多个参与方在保证各自原始数据不出数据方定义的可信域前提下，以保护隐私数据的方式交互中间数据，从而协作完成人工智能、机器学习任务的方法。

[来源：GB/T 46284—2025, 3.1]

3.7.14

隐私计算 privacy computing

面向隐私信息全生命周期保护的计算理论、方法和技术，涵盖了采集、脱敏、存储、使用、交换、删除、存证与取证等全生命周期过程的所有计算操作，包含处理视频、音频、图像、图形、文字、数值、泛在网络行为信息流等信息时，对所涉及的隐私信息进行描述、度量、控制、脱敏、使用、评价、删除等处理。

3.7.15

隐私保护计算 privacy-preserving computation

在保证数据提供方不泄露原始数据的前提下，对数据进行分析计算的一类信息技术，保障数据在产生、存储、计算、应用、销毁等数据流转全过程的各个环节中“可用不可见”。隐私保护计算的常用技术方案有安全多方计算、联邦学习、可信执行环境、密态计算等。常用的底层技术有混淆电路、不经意传输、秘密分享、同态加密等。

3.7.16

密态计算 cryptographical computation

综合运用密码学、可信硬件和系统安全的隐私保护计算技术，在加密或非透明状态下直接对数据进行运算的技术。其计算过程实现数据可用不可见，计算结果能够保持密态化，以支持构建复杂组合计算，实现计算全链路保障，防止数据泄漏和滥用。

3.7.17

差分隐私 differential privacy

一种形式化隐私度量模型，确保无论数据集中是否包含特定个体的数据，统计分析结果的概率分布差异都不会超过一个预先设定的值。

3.7.18

数据沙箱 data sandbox

一种可信管控技术，通过构建一个应用层隔离环境，允许数据使用方在安全和受控的区域内对数据进行分析处理。

参 考 文 献

- [1] 《中华人民共和国数据安全法》
- [2] 《中华人民共和国个人信息保护法》
- [3] 《数据领域常用名词解释》（第一批） 国家数据局
- [4] 《数据领域常用名词解释》（第二批） 国家数据局
- [5] 《计算机科学技术名词》（第三版） 全国科学技术名词审定委员会
- [6] 术语在线 全国科学技术名词审定委员会
- [7] GB/T 33745—2025 物联网 术语
- [8] GB/T 35295—2017 信息技术 大数据 术语
- [9] GB/T 25069—2022 信息安全技术 术语

索 引

汉语拼音索引

A		个人信息..... 3. 3. 8
安全多方计算..... 3. 7. 12		个人信息数据..... 3. 3. 9
B		公共数据..... 3. 3. 3
半结构化数据..... 3. 3. 12		公共数据资源..... 3. 3. 4
包容性..... 3. 1. 31		公共数据资源登记主体..... 3. 5. 10
标识符..... 3. 4. 15		公共数据资源授权运营..... 3. 5. 11
C		H
差分隐私..... 3. 7. 17		湖仓一体..... 3. 2. 6
产业互联网..... 3. 2. 22		J
产业数字化..... 3. 1. 19		机器学习..... 3. 4. 27
城市全域数字化转型..... 3. 6. 6		及时性..... 3. 1. 28
重标识..... 3. 7. 9		交易笔数 3. 5. 21
词元..... 3. 1. 32		接入连接器..... 3. 4. 19
D		结构化数据..... 3. 3. 11
大模型..... 3. 4. 29		静态数据 3. 3. 21
大数据..... 3. 3. 15		K
第三方专业服务机构..... 3. 5. 17		开放数据集 3. 3. 28
“东数西算”工程..... 3. 2. 16		科学数据..... 3. 3. 6
动态数据..... 3. 3. 20		可靠性..... 3. 1. 27
多模态数据..... 3. 3. 14		可信身份凭证 3. 7. 4
F		可信数据空间..... 3. 2. 7
非结构化数据..... 3. 3. 13		可信执行环境 3. 7. 3
分布式计算..... 3. 2. 10		可追溯性..... 3. 1. 30
分布式文件系统..... 3. 2. 11		L
G		冷数据..... 3. 3. 24
高速数据网..... 3. 2. 19		联邦学习..... 3. 7. 13
高质量数据集..... 3. 3. 27		M
个人数据..... 3. 3. 9		密态计算 3. 7. 16

敏感个人信息..... 3.3.10

N

匿名化..... 3.7.8

Q

企业数据..... 3.3.7

区块链..... 3.2.25

去标识化..... 3.7.7

全国一体化算力网..... 3.2.17

R

热数据..... 3.3.22

S

失效身份凭证..... 3.7.5

数场..... 3.2.23

数据..... 3.1.1

数据安全..... 3.1.20

数据安全技术..... 3.7.1

数据安全能力3.7.2

数据标签..... 3.4.14

数据标识3.4.12

数据标注3.4.13

数据标注产业3.6.2

数据标注人员3.6.3

数据标注员3.6.3

数据标准3.1.9

数据仓库3.2.4

数据产品3.1.14

数据产权3.5.3

数据产权登记.....3.5.7

数据产权登记机构3.5.8

数据产权登记平台..... 3.5.9

数据产业3.1.16

数据持有权..... 3.5.4

数据处理3.4.1

数据传输3.4.20

数据存储3.2.2

数据分级3.3.2

数据分类..... 3.3.1

数据分析..... 3.4.6

数据服务..... 3.1.15

数据供方3.5.18

数据供给方3.5.18

数据共享3.1.11

数据管理..... 3.1.8

数据合规3.5.2

数据湖3.2.5

数据汇聚3.4.3

数据基础设施3.1.21

数据集..... 3.3.26

数据计算3.2.8

数据加工..... 3.4.5

数据价值..... 3.1.3

数据交付..... 3.5.23

数据交付平台3.5.24

数据交互3.4.23

数据交换3.4.21

数据交换模型3.4.22

数据交易..... 3.5.12

数据交易撮合..... 3.5.22

数据交易额..... 3.5.20

数据交易所..... 3.5.14

数据交易中心..... 3.5.14

数据接入3.4.18

数据经营权3.5.6

数据开放..... 3.1.10

数据科学..... 3.1.2

数据可视化..... 3.4.11

数据跨境流出3.5.26

数据跨境流入.....	3. 5. 25
数据利用	3. 1. 13
数据流通	3. 1. 12
数据流通服务机构.....	3. 5. 13
数据流通服务平台企业	3. 5. 15
数据流通环境	3. 5. 1
数据目录	3. 3. 25
数据清洗	3. 4. 4
数据融合.....	3. 4. 8
数据沙箱	3. 7. 18
数据商.....	3. 5. 16
数据生命周期	3. 1. 22
数据使用权.....	3. 5. 5
数据收集.....	3. 4. 2
数据特征	3. 4. 9
数据特征层次	3. 4. 10
数据提供方.....	3. 5. 18
数据脱敏.....	3. 7. 6
数据挖掘	3. 4. 7
数据销毁	3. 4. 26
数据需方.....	3. 5. 19
数据需求方.....	3. 5. 19
数据要素	3. 1. 5
数据要素市场化配置.....	3. 6. 4
数据应用.....	3. 4. 24
数据元件.....	3. 2. 24
数据运营	3. 4. 25
数据质量.....	3. 1. 24
数据治理	3. 1. 7
数据中心	3. 2. 1
数据资产	3. 1. 6
数据资源.....	3. 1. 4
数联网.....	3. 2. 18
数字产业化	3. 1. 18

数字产业集群.....	3. 6. 1
数字经济.....	3. 1. 17
数字连续性.....	3. 1. 23
数字孪生	3. 6. 8
数字消费	3. 6. 9
算力.....	3. 2. 12
算力池化	3. 2. 14
算力调度.....	3. 2. 15
算力规模.....	3. 2. 13

W

完整性.....	3. 1. 25
温数据	3. 3. 23
物联感知数据.....	3. 2. 21
物联网	3. 2. 20

Y

衍生数据	3. 3. 19
一致性	3. 1. 29
一致性验证.....	3. 7. 10
隐私保护计算.....	3. 7. 15
隐私计算	3. 7. 14
应用场景.....	3. 6. 5
用户授权令牌.....	3. 7. 11
元数据.....	3. 3. 17
原始数据	3. 3. 18
云存储.....	3. 2. 3
云计算.....	3. 2. 9

Z

真实性.....	3. 1. 26
政务数据	3. 3. 5
知识图谱.....	3. 4. 28
直接标识符	3. 4. 16
智慧城市.....	3. 6. 7
主数据	3. 3. 16
准标识符	3. 4. 17

英文对应词索引

A		
anonymization.....	3.7.8	access connector..... 3.4.19
application scenario.....	3.6.5	data annotation..... 3.4.13
authenticity.....	3.1.26	data annotation industry..... 3.6.2
authorized operation of public data resources.....	3.5.11	data application..... 3.4.24
B		data assets..... 3.1.6
big data.....	3.3.15	data at rest..... 3.3.21
blockchain.....	3.2.25	data catalogue..... 3.3.25
C		data center..... 3.2.1
citywide digital transformation.....	3.6.6	data characteristic..... 3.4.9
cloud computing.....	3.2.9	data characteristic hierarchy..... 3.4.10
cloud storage.....	3.2.3	data circulating environment..... 3.5.1
cold data.....	3.3.24	data circulation..... 3.1.12
completeness.....	3.1.25	data circulation service agency..... 3.5.13
computility.....	3.2.12	data circulation service platform enterprise..... 3.5.15
computility pooling.....	3.2.14	data classification..... 3.3.1
computility scale.....	3.2.13	data cleaning..... 3.4.4
computility scheduling.....	3.2.15	data collection..... 3.4.2
computing power.....	3.2.12	data compliance..... 3.5.2
computing power pooling.....	3.2.14	data component..... 3.2.24
computing power scale.....	3.2.13	data computing..... 3.2.8
computing power scheduling.....	3.2.15	data convergence..... 3.4.3
consistency.....	3.1.29	data delivery..... 3.5.23
consistency verification.....	3.7.10	data delivery platform..... 3.5.24
cross-border data inflow.....	3.5.25	data demander..... 3.5.19
cross-border data outflow.....	3.5.26	data desensitization..... 3.7.6
cryptographical computation.....	3.7.16	data destruction..... 3.4.26
D		data exchange..... 3.4.21
data.....	3.1.1	data exchange center..... 3.5.14
data access.....	3.4.18	data exchange model..... 3.4.22
data analysis.....	3.4.6	data factor..... 3.1.5
		data field..... 3.2.23

data fusion.....	3.4.8	data service.....	3.1.15
data governance.....	3.1.7	data set.....	3.3.26
data grading.....	3.3.2	data sharing.....	3.1.11
data handling	3.4.1	data standard.....	3.1.9
data holding rights.....	3.5.4	data storage.....	3.2.2
data identifier.....	3.4.12	data supplier.....	3.5.18
data in motion.....	3.3.20	data trading.....	3.5.12
data industry.....	3.1.16	data trading matching.....	3.5.22
data infrastructure.....	3.1.21	data transaction volume.....	3.5.20
data interaction.....	3.4.23	data transmission.....	3.4.20
data label.....	3.4.14	data use.....	3.1.13
data labeler.....	3.6.3	data use rights.....	3.5.5
data labeling.....	3.4.13	data value.....	3.1.3
data lake.....	3.2.5	data vendor.....	3.5.16
data lakehouse.....	3.2.6	data visualization.....	3.4.11
data lifecycle.....	3.1.22	data warehouse.....	3.2.4
data management.....	3.1.8	de-identification.....	3.7.7
data mining.....	3.4.7	derivative data.....	3.3.19
data opening.....	3.1.10	derived data.....	3.3.19
data operation.....	3.4.25	differential privacy.....	3.7.17
data operation rights.....	3.5.6	digital consumption.....	3.6.9
data processing.....	3.4.5	digital continuity.....	3.1.23
data product.....	3.1.14	digital economy.....	3.1.17
data property rights.....	3.5.3	digital industrialization.....	3.1.18
data property rights registration.....	3.5.7	digital industry cluster.....	3.6.1
data property rights registration authority.....	3.5.8	digital twin.....	3.6.8
data property rights registration system.....	3.5.9	direct identifier.....	3.4.16
data quality.....	3.1.24	distributed computing.....	3.2.10
data resources.....	3.1.4	distributed filesystem.....	3.2.11
data sandbox.....	3.7.18		
data science.....	3.1.2	E	
data security.....	3.1.20	East Data West Computing Project.....	3.2.16
data security capability.....	3.7.2	enterprise data.....	3.3.7
data security technology.....	3.7.1	F	
		federated learning.....	3.7.13

G	government data..... 3.3.5	privacy computing..... 3.7.14
H	high-quality dataset..... 3.3.27	privacy-preserving computation..... 3.7.15
	high speed data network..... 3.2.19	public data..... 3.3.3
	hot data..... 3.3.22	public data resources..... 3.3.4
I	identifier..... 3.4.15	public data resources registration entity..... 3.5.10
	inclusiveness..... 3.1.31	Q
	industrial digitization..... 3.1.19	quasi-identifier..... 3.4.17
	industrial internet..... 3.2.22	R
	internet of data..... 3.2.18	raw data..... 3.3.18
	internet of things..... 3.2.20	reidentification..... 3.7.9
	invalid credential..... 3.7.5	reliability..... 3.1.27
	IoT..... 3.2.20	S
	IoT sensing data..... 3.2.21	scientific data..... 3.3.6
K	knowledge graph..... 3.4.28	secure multi-party computation..... 3.7.12
L	large model..... 3.4.29	semi-structured data..... 3.3.12
M	machine learning..... 3.4.27	sensitive personal information..... 3.3.10
	market-oriented allocation of data factor..... 3.6.4	smart city..... 3.6.7
	master data..... 3.3.16	SMPC..... 3.7.12
	metadata..... 3.3.17	structured data..... 3.3.11
	multi-modal data..... 3.3.14	T
N	national integrated computility network..... 3.2.17	third-party professional service organization..... 3.5.17
	national integrated computing power network.... 3.2.17	timeliness..... 3.1.28
O	open dataset..... 3.3.28	token..... 3.1.32
P	personal information..... 3.3.8	traceability..... 3.1.30
	personal information data..... 3.3.9	transaction count..... 3.5.21
		trustworthy data space..... 3.2.7
		trustworthy execution environment..... 3.7.3
		trustworthy identity credential..... 3.7.4
		U
		unstructured data..... 3.3.13
		user authorization token..... 3.7.11
		W
		warm data..... 3.3.23

