

附件

数据领域常用名词解释（第一批）

1.数据，是指任何以电子或其他方式对信息的记录。数据在不同视角下被称为原始数据、衍生数据、数据资源、数据产品和服务、数据资产、数据要素等。

2.原始数据，是指初次产生或源头收集的、未经加工处理的数据。

3.数据资源，是指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。

4.数据要素，是指投入到生产经营活动、参与价值创造的数据资源。

5.数据产品和服务，是指基于数据加工形成的，可满足特定需求的数据加工品和数据服务。

6.数据资产，是指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。

7.数据要素市场化配置，是指通过市场机制来配置数据这一新型生产要素，旨在建立一个更加开放、安全和高效的数据流通环境，不断释放数据要素价值。

8.数据处理，包括数据的收集、存储、使用、加工、传输、提

供、公开等。

9.数据处理器，是指在数据处理活动中自主决定处理目的和处理方式的个人或者组织。

10.受托数据处理器，是指接受他人委托处理数据的个人或者组织。

11.数据流通，是指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。

12.数据交易，是指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。

13.数据治理，是指提升数据的质量、安全、合规性，推动数据有效利用的过程，包含组织数据治理、行业数据治理、社会数据治理等。

14.数据安全，是指通过采取必要措施，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。

15.公共数据，是指各级党政机关、企事业单位依法履职或提供公共服务过程中产生的数据。

16.数字产业化，是指移动通信、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。

17.产业数字化，是指传统的农业、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。

18.数字经济高质量发展，是指围绕加快培育新质生产力，以数据要素市场化配置改革为主线，通过协同完善数据基础制度和数字基础设施、全面推进数字技术和实体经济深度融合、持续提升数字经济治理能力和国际合作水平，实现做强做优做大目标的数字经济发展新阶段。

19.数字消费，是指数字技术、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。

20.产业互联网，是指利用数字技术、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。

21.城市全域数字化转型，是指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。

22.“东数西算”工程，是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，比如人工智能模型训练推理、机器学习等业务场景，可以通过“东数西算”的方式让东部业务向西部风光水电丰富的区域迁移，实现东西部协同发展。加快推动“东数西算”工程建设，将有效激发数据要素创新活力，加速数字产业化和产业数字化进程，催生新技术、新产业、新业态、新模式，支撑经济高质量发展。

23.高速数据网，是指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。

24.全国一体化算力网，是指以信息技术为载体，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。

25.元数据，是定义和描述特定数据的数据，它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。

26.结构化数据，是指一种数据表示形式，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。

27.半结构化数据，是指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。

28.非结构化数据，是指不具有预定义模型或未以预定义方式组织的数据。

29.数据分析，是指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。

30.数据挖掘，是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。

31.数据可视化，是指通过统计图表、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。

32.数据仓库，是指在数据准备之后用于永久性存储数据的数据库。

33.数据湖，是指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。

34.湖仓一体，是指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。

35.隐私保护计算，是指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。

36.安全多方计算，是指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。

37.联邦学习，是指一种多个参与方在保证各自原始私有数据不出数据方定义的可信域的前提下，以保护隐私数据的方式交换中间计算结果，从而协作完成某项机器学习任务的模式。

38.可信执行环境，是指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。

39.密态计算，是指通过综合利用密码学、可信硬件和系统安全相关技术，实现计算过程数据可用不可见，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。

40.区块链，是分布式网络、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过程中的信任和安全问题。