

数据要素场技术体系及工程实践

吴曼青¹, 洪日昌^{2*}, 王佐成³, 林传文³, 马韵洁³, 郭嘉丰⁴, 吴乐³, 范举⁵, 张兰³, 王翔³

(1. 中国工程院, 北京 100088; 2. 合肥工业大学计算机与信息学院, 合肥 230009; 3. 数据空间研究院, 合肥 230071; 4. 中国科学院计算技术研究所, 北京 100190; 5. 中国人民大学信息学院, 北京 100872)

摘要: 将数据作为新的生产要素, 是我国在精准把握和研判全球科技发展规律下提出的重大理论创新。以数据要素市场化配置改革为主线, 培育全国一体化数据市场, 促进数据要素开发利用, 是我国数据要素创新发展的总体纲领。本文围绕数据要素市场化配置改革, 聚焦推动数据要素流通和数据要素价值释放, 提出探索数据要素价值时空分布的内在机理即数据场基础理论, 探讨了在深入研究数据场基础理论的同时, 构建涵盖数据要素流通全生命周期的数据要素场技术体系, 具体包括跨域数据管理技术、数据件封装技术、低熵化流通技术、穿透式安全技术和聚变式处理技术。同时, 分析了数据要素场在卫生健康场景中的工程实践案例, 提出了数据要素场的创新应用场景和工程实践范式, 展望了数据场基础理论和数据要素场关键技术、工程实践、生态构建方面的前景, 旨在为数据场的发展提供理论基础和实践指导, 推动数字经济和社会治理的现代化。

关键词: 数据场; 跨域数据管理; 数据件封装; 低熵化流通; 穿透式安全; 聚变式处理技术

中图分类号: TP315 文献标识码: A

Technology System and Engineering Practice of Data Field

Wu Manqing¹, Hong Richang^{2*}, Wang Zuocheng³, Lin Chuanwen³, Ma Yunjie³,
Guo Jiafeng⁴, Wu Le³, Fan Ju⁵, Zhang Lan³, Wang Xiang³

(1. Chinese Academy of Engineering, Beijing 100088, China; 2. School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, China; 3. Institute of Data Space, Hefei 230071, China; 4. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China; 5. School of Information, Renmin University of China, Beijing 100872, China)

Abstract: The concept of data as a new factor of production is a significant theoretical innovation put forward by China, based on a precise understanding and prediction of global technological development trends. With the market-oriented allocation reform of data elements as the main line, cultivating a national integrated data market to promote data utilization is the general program for the innovative development of China's data elements. It introduces the fundamental theory of a data field and examines key technologies such as cross-domain data management, data encapsulation, low-entropy data circulation, penetration security, and fusion processing technology. Moreover, the study analyzes engineering practice cases of data fields in health scenarios, proposes innovative application scenarios and engineering practice paradigms for data fields, and looks forward to the prospects in the fundamental theory, key technologies, engineering practices, and ecological construction of data fields, aiming to provide theoretical and practice guidance for data field development and support the modernization of the digital economy and social governance.

收稿日期: 2024-11-02; 修回日期: 2024-12-02

通讯作者: *洪日昌, 合肥工业大学计算机与信息学院教授, 主要从事人工智能相关研究; E-mail: hongrc@hfut.edu.cn

资助项目: 中国工程院咨询项目“国家数据空间发展路径与技术体系研究”(2024-XBZD-05)

本刊网址: ssc.ae.engineering.org.cn

Keywords: data field; cross-domain data management; data encapsulation; low-entropy circulation; penetration security; fusion processing technology

一、前言

数据作为一种新兴的生产要素，在推动社会进步、促进经济发展以及保障国家安全等方面的作用日益凸显，成为国家战略资源的重要组成部分。我国高度重视数据要素产业的发展，先后发布了《“十四五”数字经济发展规划》^[1]、《关于构建数据基础制度更好发挥数据要素作用的意见》^[2]、《关于加快公共数据资源开发利用的意见》^[3]、《关于完善数据流通安全治理 更好促进数据要素市场化价值化的实施方案》^[4]等政策文件。全国各地积极响应国家号召，北京、上海、广州、深圳、杭州等地纷纷设立数据交易所，纷纷出台政策文件加快探索公共数据的授权运营模式，积极推动数据要素的流通利用与市场化交易。加强数据要素基础理论研究和数据流通利用关键技术，不仅有助于夯实政策实施的理论基础，还能为数据交易定价、流通机制及安全保障等核心环节提供技术支撑，确保政策措施的有效落地，进一步提升数据要素市场化配置水平。

在全球范围内，数据要素的发展同样引起广泛关注，各国积极制定战略规划以释放数据的潜在价值。美国通过《开放政府数据法案》^[5]《联邦数据战略》^[6]等政策，加强数据管理和利用，提升在全球数字经济中的竞争力。欧盟及欧洲国家分别发布了《2030 数字罗盘》^[7]《英国数字战略》^[8]《数字爱尔兰框架》^[9]等规划，激发数据要素的活力，进一步完善数据流通的规则。

然而，在数据要素的研究与实际应用中，我们仍面临着对数据本质属性理解不足和基础理论框架不完善的挑战^[10,11]。为此，本文提出数据场理论的设计与机理，以深化对数据要素的理解，并构建一个全面且成熟的数据场技术体系，简要阐明该技术体系的工程实践和在医疗健康领域的工程实践案例，以期支撑数据要素的深度应用和高效发展。这一体系不仅填补了现有理论与实践中的空白，更为推动我国在全球数据要素战略中占据重要地位、巩固我国在数字经济领域的领导地位提供了有力支持。

二、数据场的机理与特征

在物理学中，场是物质存在的一种形式，与实物一样具有能量、质量和动量，场量被定义为针对空间和时间的函数。这一概念涵盖了电磁场、引力场、原子力场等多种类型。具体而言，一个带有质量的物体会在其周围形成引力场，引力场在空间的分布本质，即在实体空间中，任何一个空间位置在任意时刻都有一个引力势；粒子的电荷属性和自旋属性产生了电磁场，电磁场在空间分布的本质，即在任何一个空间位置，在任意时刻都有一个电磁势，如图1所示。当一个物体被置于引力场或者电磁场中，作用于物体的力等于该物体的质量或者电荷乘以物体所在位置的场矢量数值，这种简洁而深刻的数学关系，反映了物理规律的普适性。那么在数据流通过程中，数据在数据空间内的流通是否形成了数据场，数据场内数据受到的相互作用是否也

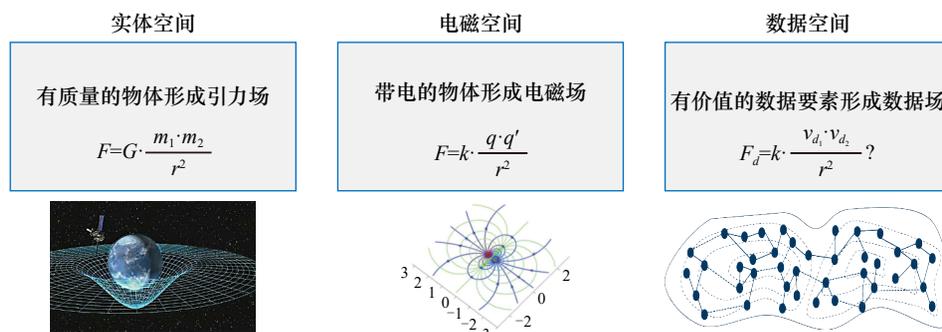


图1 数据场形成示意图

能表达为数据的某一内禀属性与场矢量的乘积，则需要进一步研究。

本文认为，数据场理论也是描述物理量之间的关系及其空间分布与时间变化的理论。力的本质在于传递，在引力场中，我们感受到地球的引力，本质上是因为我们接收到了“引力子”；在电磁场中，电荷之间互相排斥和吸引，本质是因为电荷之间交换了光子；而数据天然带有交互和传递的属性，因此数据可以和基本粒子对比，而数据场即是数据传递作用力的介质。

（一）数据场的机理

在数据空间中，数据要素历经生产、分配、交换和消费等一系列复杂而精密的社会再生产过程，最终促成价值的释放。这一过程蕴含着推动数据要素价值释放的深层次规律和机制。本文认为，数据场是数据要素价值与相互作用在时空上的分布，刻画了数据要素在数据空间中运动的基本规律。在数据场的作用下，无序的数据要素有序地流通，有序的数据要素流通持续地创造价值。

数据场理论的提出旨在通过精细描述与动力学测度数据空间中的要素及其互动情况，揭示数据价值流通机理，设计数据流通机制，突破核心技术瓶颈，解决流通使用中的规律、权责、安全、技术等问题，有效应对数据商品化、要素化、市场化改革中的确权、定价、互信、入场、监管难题，为设计与实现数据要素在数据空间中高效流通的手段和方法提供理论指导。

（二）数据场的特征

本文提出数据场的3个核心特征，这些特征共同构成了数据场的基本属性。

1. 价值链接性

数据场贯穿数据要素的生产开发、流通共享、价值实现及安全保障等全生命周期各环节，通过构建完整的价值链条，实现数据要素的高效流转与价值增值。

2. 动态流通性

数据场具有独特的时空动态特性，能够促进数据要素在不同时间、空间维度上的高效流动，保证数据价值的及时释放与传递。

3. 协同互联性

在数据场中，数据要素不再是孤立的存在，而是通过相互作用和联动，形成一个具有高度协同效应的整体。数据要素通过相互作用和联动形成网络化结构，突破单一“数据孤岛”的局限，产生整体大于部分之和的协同效应。

通过数据场的构建，数据要素得以在一个高效、安全的环境中流通和转化，促进数据的价值创造。这种价值创造不仅体现在经济效益上，还体现在社会治理、公共服务、科研创新等多个领域。数据场通过优化数据流通过程、标准化数据处理、保障数据安全等手段，为数据价值的释放提供了坚实的技术基础和保障机制。

（三）数据价值的释放模式

本文提出数据要素价值的释放模式主要包括关联释放、聚变释放和倍增释放3种模式，这些模式反映了数据要素在不同场景下的价值创造机制和路径。通过这些释放模式，可以更深入地理解数据场如何通过数据要素的高效流通，实现数据价值的最大化。

1. 关联释放模式

数据要素是一种新兴的数字化生产要素，其特点在于其内蕴的语义与结构。通过构建多样化的连接关系，数据要素对象间可相互作用，催生新的价值。借助数据开放、联盟、市场等平台，能促成跨行业、跨领域、跨层级、跨组织的数据互联互通与互惠合作，推动数据价值的流动、协同与优化。这一过程被称为关联释放模式，其核心在于通过关联技术将原本分散稀疏的数据要素建立为广谱关联网络，并利用网络化效应实现数据价值的最大化。

2. 聚变释放模式

在信息空间中，信息的价值涌现表现为“裂变”模式：信息借由互联网络广泛传播，连接节点越多则用户触及面越大，信息价值随之提升。梅特卡夫定律揭示，网络信息空间的价值与用户数量的平方成正比。数据空间中的数据要素价值涌现则类似“聚变”模式，即通过汇聚、融合及深度处理海量数据，提炼知识，实现数据价值的涌现。以大模型为例，数据汇聚与提炼的规模和多样性决定价值收益的高低。

3. 倍增释放模式

数据要素的倍增释放模式，即通过与其他生产要素的结合，产生超越单一要素的价值，推动数字经济与实体经济的深度融合，实现高质量发展。其核心在于数据要素的协同、复用与融合效应，以低成本实现量变至质变的转化。此模式作为数据驱动决策与创新的关键，能够激励组织发掘数据潜力，驱动组织和社会进步，成为行业和经济转型的强劲动力。

三、数据要素场的技术体系

在初步探讨数据场的机理与特征后，本文进一步明确数据场作为数据要素流通基本物理规律的假定和设想。当前，以新技术推动数据要素流通并释放价值形成新质生产力已迫在眉睫。然而，目前尚存在诸多挑战，有待进一步深入探索，因此需要同步开展如何通过技术手段保障数据要素高效流通的研究，即数据要素场技术体系构建。二者的关系如下：数据场理论将为设计数据要素流通和价值释放的手段和方法提供理论指导；反之，攻克数据要素场面临的技术挑战也将为进一步明确数据场理论提供支撑。简言之，数据场理论与数据要素场体系两者之间相辅相成，两者的发展呈螺旋式上升。

本文认为，数据要素场技术体系是支撑跨域数据高效流通的一类新型基础设施，提供跨域数据的统一语义、统一封装、可控流通、高效分发等能力，支撑数据要素有序流通和持续释放价值。数据要素场技术体系针对数据基础设施的建设需求，将五大关键技术体系有机融合，综合运用关键技术解决数据在跨域流通、标准化处理和安全保障中的诸多挑战，为数据要素的高效流通和使用提供了系统化、标准化、智能化支撑。数据要素场技术体系包含以下五大核心技术，如图2所示。

（一）数据件封装技术

数据件封装技术是一种新的数据要素标准化抽象方法，旨在支撑数据要素的高效流通与汇聚使用。数据件作为数据要素流通使用的基本单位，是对数据要素进行标准化抽象后建立的结构化对象，基于一组标准、协议与机制设计，具有可寻址、可交换、可操作、可管控的特性。数据件封装技术包

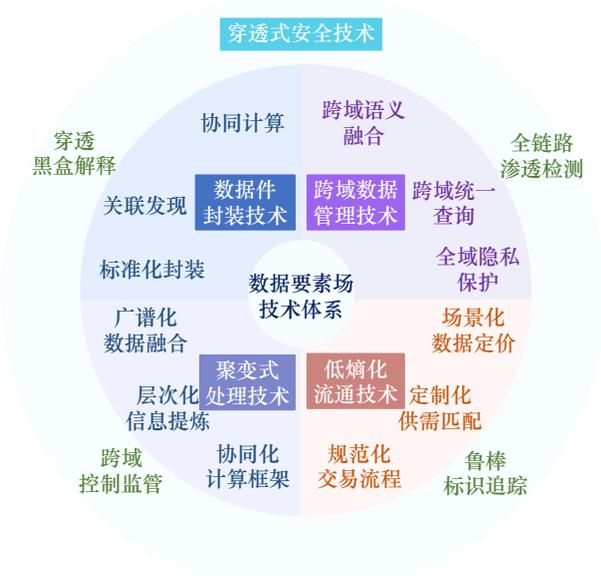


图2 数据要素场技术体系示意图

括但不限于以下几个方面。

1. 高可用的数据件构造方法

针对现有数据标准化方法面临数据内容质量低下导致利用价值低、标识缺乏语义难以高效检索、操作方法容易暴露隐私信息等挑战，围绕数据要素多源异构与统一表征之间的矛盾，研究高质量数据要素基础元件构造方法，包括语义对齐标识方法、效用提升蒸馏方法、敏感约束接口生成方法，实现多元异质数据的统一表征，完成数据要素的标准化构造，解耦数据源和数据应用，推动数据要素高效流通与便捷使用。高可用的数据件构造方法，是数据要素灵活适用于复杂多样应用场景的基础。笔者提出的数据件基本结构如图3所示。

2. 去中心化的数据件关联发现方法

针对当前数据要素关联发现基于简单模式匹配，导致关联通用性弱、组织效率低、发现精度差等方面的问题，针对数据要素分布无序与精准发现

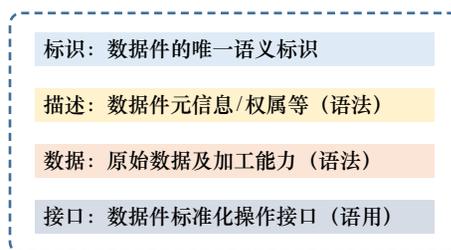


图3 数据件基本结构图

之间的矛盾, 研究去中心化的数据件关联发现方法, 包括广谱关联方法、分布式图网络方法以及结构化匹配发现方法等, 实现通用性强、效率高、精度优的数据件关联发现, 打通广谱表征、分布组织与匹配发现, 进而提升基础元件关联发现方法在真实应用中的可用性。去中心化的数据件关联发现方法是保障数据要素面向应用高效查找和灵活运用的核心。

3. 面向广域分散数据件的协同计算技术

针对数据协同计算所面临的数据广域分散、体量庞大、时效性要求高等挑战, 以及数据要素广域分散与高效运用之间的矛盾, 研究面向广域分散数据件的协同计算关键技术, 包括资源自适应编排技术、自组织协同加速技术和垂直优化计算技术。其中, 基于自适应资源特性的广域编排技术可实现最优资源调度; 自组织的联合传输加速技术, 降低传输响应时间; 基于垂直优化等计算加速方法提升数据加工效率, 从而满足数据件高效的广域协同计算要求。面向广域分散数据件的协同计算技术是实现数据要素的高效协同加工、使用, 更好地融合、释放数据要素价值的关键。

(二) 跨域数据管理技术

跨域数据管理建立在数据要素标准化抽象的基础上, 旨在对分散于不同域(如部门、层级等)的数据进行统一管理, 构建语义可融合、数据可保护、查询可统一的高效数据管理系统, 实现数据在跨域间进行高效、安全的共享流通, 从而最大限度地释放数据要素价值。本文围绕以下3个方面具体介绍跨域数据管理所包含的关键技术。

1. 跨域异质数据的语义融合技术

在数据的供给侧, 需要解决的核心问题是跨域多方的数据异质, 实现“找得准”。具体而言, 跨域开放环境带来的数据结构不统一、数据语义难对齐、数据来源动态多变等挑战, 突出体现为跨域数据异质与统一语义表示之间的矛盾, 已成为数据管理领域的研究热点^[12]。围绕跨域开放环境带来的新挑战, 数据要素场重点突破跨域异质数据的语义融合技术, 为跨域数据生成统一的数据模式, 实现跨域数据模式的统一语义表示, 支持跨域数据模式的高效语义匹配, 从而满足跨域数据管理对于统一语义表示与高效数据发现的核心需求。跨域异质数据

的语义融合是支撑跨域数据管理的基础^[13]。

2. 跨域协同隐私保护方法

在数据的供需间, 需要解决的核心问题是跨域数据保护的复杂协同, 实现“流得通”。具体而言, 跨域数据安全面临规则多样异质、跨域多方隐私保护协同困难等挑战, 突出体现为跨域情况下的隐私规则多样与跨域隐私保护之间的矛盾, 现有跨域数据流通技术在性能方面尚存在明显不足^[14]。为此, 本文提出研究域隐私保护规则冲突避免的协同隐私保护方案、设计跨域隐私算子的协同计算方法, 支持跨域隐私数据的可控生成, 从而形成一套冲突避免、协作优化的跨域协同隐私保护方法。跨域数据保护的协同计算可以显著提升跨域数据安全共享流通的性能。

3. 跨域数据查询的性能优化技术

在数据的需求侧, 需要解决的核心问题是跨域多方的资源异构, 实现“查得快”。跨域数据源自治给跨域查询处理带来的底层数据库类型多样、数据资源异构、查询动态变化等难题, 突出体现为跨域资源异构与高效数据查询之间的矛盾, 给现有的跨域查询研究带来了新的挑战^[14,15]。为此, 本文提出研究跨域异构资源感知的跨域查询性能优化技术, 支持场景适应的跨域查询翻译与动态查询调度, 从而满足跨域查询的高效性要求。跨域数据查询的性能优化是支撑跨域数据共享流通高效性的关键。

(三) 聚变式处理技术

聚变式处理技术旨在通过融合多种数据要素理论与技术, 推动数据要素在复杂的数据竞价交易环节中数据价值的估计与衡量, 充分挖掘数据价值。该技术框架包括广谱化数据融合、层级化信息提炼和协同化计算框架这三大核心。广谱化数据融合为数据要素建立了坚实的价值基础, 层级化信息提炼确保了竞价机制的科学性与精确性, 而协同化计算框架则通过智能化的演化计算, 保证了数据要素流通的高效性。

1. 广谱化数据融合

广谱化数据融合旨在通过将不同类型的数据要素与多样化的应用场景进行有机结合, 形成一个广泛而灵活的数据估值体系。数据要素的价值随着应用场景、供需关系、买卖双方和交易时效的变化而

变化，这导致了数据要素价值在流通交易中的多变性和不确定性。一方面，数据要素相比传统商品具有外部性、时效性和差异性特征；另一方面，大模型领域的研究表明，大语言模型所具有的世界知识使其具有强大的泛化与适应能力。结合传统机器学习模型对数据要素特征进行可解释建模，以及大模型在不同应用场景和供需关系中的适应能力，可以将不同领域和渠道的数据进行整合分析，完成数据要素的广谱化数据融合，进而开发出适应复杂多样市场环境的估值技术，为数据要素建立坚实的价值基础。

2. 层级化信息提炼

在复杂的数据竞价交易环境中，信息不对称、假名竞价攻击和策略性误报等问题常常导致交易机制的失效^[15]。通过引入层级化信息提炼技术，可以分阶段分析买家的决策行为和交易机制，在每个层次上进行精准的调整和优化，以提炼出稳定的策略组合与最优的竞价机制。具体而言，该技术借助分阶段的博弈模型^[16]，对买家行为和竞价机制进行精确建模与验证：通过使用大模型工具模拟真实买家的决策过程，为机制分析提供更精准的数据和决策支持；并进一步依照买家均衡结果验证并调整竞价机制、分析其相互影响因素，从而将信息处理分层次进行，从宏观市场动态到微观个体决策都进行深度提炼，确保每个层级的信息都能有效支持决策。

3. 协同化计算框架

随着数据要素应用场景的不断演化，市场参与者之间的博弈模式也在变化。新兴场景极大地扩展了数据的类型和用途，也相应提高了数据价值评估和交易的复杂度。更重要的是，为历史场景设计的竞价机制很难适应新兴场景的需求和特点，导致历史上的均衡状态在新场景下无法提供有效的指导和参考，从而使新场景中的均衡状态变得难以预测和监管。协同化计算框架通过构建数据要素交易模拟环境，利用博弈论、智能体建模以及人机混合智能等技术，实现数据要素交易的动态均衡与持续优化^[17]。具体来说，此模拟环境在广谱化数据融合和层级化信息提炼的基础上，结合博弈论理论和大模型技术，搭建具有理论收敛保障和复杂行为模拟能力的智能体，来对数据要素竞价参与者进行准确、真实的建模。通过智能体自主竞价，模拟环境能够将推演出数据要素在真实场景下的流通过程以及最

终的竞价均衡。

（四）低熵化流通技术

低熵化流通旨在创造一个有序、高效、结构化的多源数据要素流通生态，支撑数据要素顺畅流通。在数据交易平台上，制定规范的数据交易流程范式、适应不同场景的数据定价方法以及个性化的供需匹配策略，确保数据需求方、供应方以及平台等各方的参与积极性，最大限度保证数据高效流通。低熵化流通技术包括但不限于以下几个方面：场景化数据定价、定制化供需匹配和规范化交易流程。

1. 场景化数据定价方法

场景化数据定价是指针对不同数据要素流通场景，使用差异化方法对数据价值进行衡量与价格制定。常见的要素流通场景包括数据全部可见、数据部分可见和数据完全不可见。突破数据价值的场景依赖与精准衡量之间的矛盾，解决数据对于不同任务价值贡献难度量、部分数据缺失导致数据价值难评估、数据完全无法感知带来的价值估计困难等挑战，需要建立数据和模型相结合的数据要素定价机制；提出基于数据贡献度的定价策略、基于图基础模型的多任务定价策略、基于多智能体主动协商的定价策略，覆盖数据要素全部可见、部分可见和完全不可见等多种流通场景，支撑单一供方到多供方并存的复杂数据定价需求^[18]。场景化数据定价是数据要素低熵化高效流通的基础。

2. 定制化供需匹配策略

定制化供需匹配是指在数据交易过程中，为需求方匹配个性化的供方数据，满足供需双方的功能与价格要求，从而达成数据交易。突破数据交易的多元约束与个性化供应之间的矛盾，解决供需匹配样本稀缺、样本组合价值难衡量和扰动方式与模型性能关系的不确定性，需要定制符合需求方出价的数据要素产品、挑选匹配需求方价格条件的数据要素产品定制化模型，实现定制化供需匹配策略设计；通过计算数据描述表征和用户需求表征的匹配度，设计基于功能匹配的供需撮合机制；根据卖方数据能否感知，设计基于数据价值或模型价值的匹配范式。定制化供需匹配是数据要素低熵化高效流通的关键。

3. 规范化交易流程范式

规范化交易流程是指对数据交易流程进行合规

化处理。针对数据权属难界定、数据交易难合规、数据纠纷难解决等挑战,依托功能完善的数据要素交易平台,需要建立完备的数据要素流通准则与透明公开的交易制度。在数据交易前,通过区块链技术进行数据权属认证,确保所有权明确、数据内容合规,保障交易合法合规^[19]。在数据交易中,依托正规数据交易平台确保交易透明、公正,引入第三方监管,实时监控交易过程,增强交易可信度。数据交易完成后,供需方确认合同约定,平台出具验收报告,建立纠纷解决机制,保障交易双方权益。规范化交易流程范式是数据要素低熵化高效流通的重要保障。

(五) 穿透式安全技术

安全保障是数据要素流通的前提和底线。不同于传统要素,数据具有非独占、增长速度快、复制成本低、易删除易修改、潜在价值未知、流通渠道难管控、隐私信息丰富等特性,使得数据要素流通面临诸多特有的安全挑战^[20]。

根据数据要素三次价值释放的过程和当前数据要素流通市场的交易标的,数据要素流通链路分为三类:一是原始数据流通链路,即直接将经过脱敏的一条数据或者多条数据组成的数据集/数据流进行流通;二是信息流通链路,即将原始数据经过特定计算(如安全多方计算)之后的结果进行流通;三是知识和智慧流通链路,通过人工智能算法将数据学习为有价值的知识以及模型参数(如联邦学习模型和大型语言模型)进行流通。三类链路面临不同的安全挑战,依靠不同的技术体系实现数据流通,释放数据价值。

1. 可追溯的数据安全

数据模态多样,且易复制、易修改、易传播,因此原始数据的流通链路往往面临着大量来源不明,隐私丰富的数据,这些数据通过各种传播渠道最终去向不明,对数据本身的隐私和数据应用安全带来了巨大威胁。尤其是数据具有聚合效应,当汇聚多个时空的数据,可能使得原本各自安全的数据产生严重的隐私泄露。问题在于如何平衡数据标识的鲁棒性和可区分性,以及如何建模跨时空数据融合产生的隐性隐私泄露。需要研究在原始数据流通链路中的跨域数据可信验证和追溯,包括多模态数据和人工智能模型的抗篡改鲁棒指纹和水印^[21],及

其穿透不同数据域的高效可靠追溯验证机制。对于恶意环境的多模态数据跨时空隐私安全,需要研究多模态数据的语义关联及隐私分析,以及穿透不同时空的数据关联建模和融合数据的隐私分析及脱敏技术。

2. 可验证的计算安全

为实现智能化决策和数据流通对外赋能,常常采用多方安全计算的方式实现多方之间保护隐私的联合数据分析。然而多方安全计算常常依赖半诚实假设,并且只支持一定类型的计算,面对不可信的恶意参与方,过程复杂的计算依然难以保障数据的隐私安全和结果的正确性^[22],问题在于从简单计算原语到复杂组合计算的安全性分析理论,以及如何解决计算全过程安全性和可验证性的矛盾。针对该问题,需要研究在信息流通链路中的安全、高效跨域隐私计算,包括面向恶意环境的多类型安全计算原语的高效实现、复杂安全计算原语组合的安全性分析和效率优化以及面向数据动态更新与动态计算需求的高效隐私计算理论与协议。

3. 可解释的模型安全

随着人工智能技术的突飞猛进,数据要素价值的释放越来越依赖机器学习模型^[23],尤其是深度神经网络。联邦学习、拆分学习等技术使得多方可以不传输原始数据而直接联合训练模型并进行推理。然而,近年来针对神经网络的数据重构、标签推断、后门嵌入等攻击层出不穷,导致了隐私泄露、判断错误、功能失效等后果^[24],可能造成难以弥补的经济和安全损失。这一问题在于复杂分布式神经网络的全面安全性分析和可解释理论,以及大模型的通用性和个性化之间的矛盾。通过研究穿透复杂网络模型,从输入到参数,再到输出特征和推理结果的模型安全可解释理论,结合分布式机器学习模型的不同攻击面,以及面向不同攻击面的安全攻击和防御机制,实现全面防御。如果面向大规模集成模型,则侧重穿透多异构模型的安全分析理论和隐私保护机制的研究。

通过数据件封装技术、跨域数据管理技术、聚变式处理技术、低熵化流通技术和穿透式安全技术的相互作用与有机融合,形成了一个完整的闭环体系。跨域数据管理技术提供了数据共享的基础,确保数据可以在不同域之间流通;数据件封装技术对跨域共享的数据进行标准化处理,使其便于流通和

使用；聚变式处理技术描述了复杂数据竞价交易环节中数据价值的估计与衡量；低熵化流通技术优化了标准化数据的流通过程，提高了数据流通的效率和有序性；穿透式安全技术确保数据在流通过程中的安全性，防止数据泄露和滥用^[4]。通过数据要素场技术体系五个核心技术的有机融合，实现了数据在生成、流通、交换和消费各环节的高效、安全和有序流动，充分释放了数据的价值，推动数字经济和社会治理的现代化。

四、数据要素场的工程实践

目前跨域数据流通面临语义统一难、标准化利用难、使用权益保障难、可信可控流通难、流通效率低等问题，传统数据流通技术多关注于数据流通中的某一环节，例如，在欧洲，国际数据空间协会（IDSA）以国际数据空间连接器实现数据的安全和可信交换；美国通过实施国家信息交换模型项目，解决跨域信息共享问题。然而，这些独立的技术不能满足数字时代体系化数据流通需求，需要新的工程实践体系为数据空间内的数据流通提供支撑。

基于数据场理论，在数据场的影响与作用下，实现无序的数据要素有序地流通，有序的数据要素流通持续地创造价值。针对语义、标准化、控制等国际数据流通研究重点，聚焦跨区域、跨行业、跨机构数据互联互通和互操作，通过搭建系统化、标准化、智能化的数据流通与利用基础设施，形成完善的数据要素场工程实践体系，提升数据高效流通利用和跨域融合创新水平，支撑数据要素价值关联释放、聚变释放和倍增释放。

（一）工程实践架构

数据要素场工程实践依赖于数据要素场技术体系，五大核心技术为数据高效流通提供了坚实的技术支撑和保障。数据要素场工程实践架构如图4所示，主要包括节点模块、数据连接模块、数据流通模块、数据服务模块和价值释放模块。

节点与数据连接模块，通过对数据进行接入、存证和标准化构造，实现数据的接入与标准化处理。节点模块为数据流动提供的接入口，主要依托算力网络基础设施实现数据的接入。数据连接模块，将节点接入的数据进行登记和标识，实现数据的接

入、登记、标识注册、封装、交换标准化。其中，国家数据登记存证实现数据全生命周期行为存证，登记认证数据各方的权属。国家数据统一标识实现全国数据统一标识，兼容国际主流标识体系。数据件构造功能依托数据件封装技术通过对语义、标识等标准化封装，支撑数据要素与数据主体、数据应用的解耦。国家数据交换模型建立跨域数据交换与共享统一标准，支撑“一对多、多对多”数据共享交换。穿透式安全技术通过多模态数据脱敏技术对原始数据进行脱敏处理，保证原始数据安全。

数据流通模块，通过对数据件进行管理分发、聚合控制、价值评估和供需匹配，实现数据的有序、高效流通。基于跨域数据管理技术，跨域数据管理为数据各方提供数据融合、可信溯源、行为存证等方式，实现跨区域、跨层级异构数据的有序管理。国家数据分发网络通过优化请求调度、管理监控、分发服务等分发过程，提升数据的传输效率。数据聚合通过机密计算、隐私计算、数据沙箱等聚合方式支撑数据要素发挥关联、聚变、倍增效应。使用控制通过系统级控制、应用级控制和审计全程保障数据提供者对数据的控制权。穿透式安全技术通过跨域身份认证、算法测试等技术保证数据流通过程安全。

数据服务模块，依托聚变式处理技术对复杂竞价场景下数据的价值进行合理的评估，结合低熵化流通技术，实现对数据供给方和数据需求方的供需匹配与撮合。同时，还为数据交易提供其他各类服务，如检索查询、数据计量等。

价值释放模块，基于节点模块、数据连接模块和数据流通模块，结合数据场核心技术，充分赋能社会治理、数字社会、金融、卫生健康等典型应用场景，促进数据价值释放。

在数据要素场工程实践过程中，数据要素场技术体系的关键技术保障了数据的连接、流通与价值释放。其中，数据件封装技术将跨域流通的数据进行标准化处理，使数据在不同系统和应用中具有一致性和可操作性。跨域数据管理技术确保了数据在不同区域、行业和机构间的高效流通，解决数据共享的基础问题。聚变式处理技术与低熵化流通技术通过明晰数据价值，结合供需关系，优化数据流通的路径和方式，减少了不必要的损耗和延迟，提高数据流通效率。穿透式安全技术在整个数据流通过

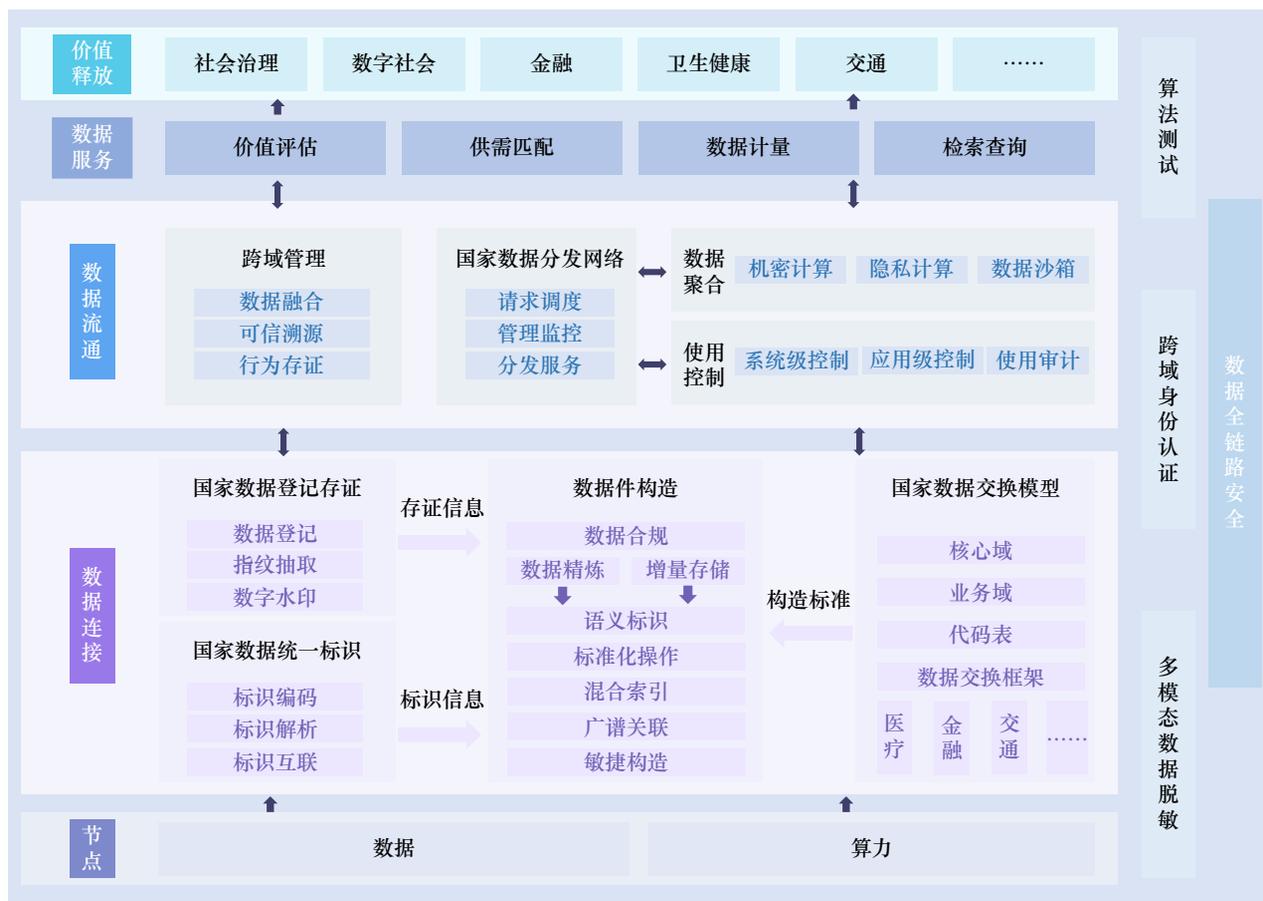


图4 数据要素场工程实践架构图

程中提供了全方位的安全保障，确保数据在不同环节中的安全性和可控性，防止数据泄露、滥用和未授权访问。

在工程实践中，数据要素场综合应用五大技术体系，成功实现了跨区域、跨行业、跨机构的数据互联互通，提升了数据的流通利用效率，促进了数据要素的高效释放和价值创造。这种工程实践不仅验证了技术体系的有效性，也展示了数据要素场在推动数字经济和社会治理现代化方面的巨大潜力。

（二）工程实践典型案例

在实践中，数据要素场的实际工程落地需要考虑更多的因素与变量。以卫生健康数据要素场的工程实践为例（见图5），分析数据要素场工程实践的总体流程以及数据要素场技术体系如何有效支撑数据要素场的工程实践。

卫生健康数据要素场全链路流通体系以数据要素的流向为基础，包括数据的采集连接、数据的标

准化、数据的聚合管理与使用及数据的场景应用。卫生健康数据通过如构建医疗大模型、疾病分析等综合应用，才能释放数据的价值。

在节点方面，卫生健康数据要素场节点模块综合利用隐私计算、数据沙箱等技术，依托算力网络基础设施，实现数据的物理汇聚与节点逻辑连接。汇聚卫生健康领域中的医院、医疗共同体、卫生监管部门、体检机构等主体的卫生健康数据，安全有效地完成数据采集任务，为数据要素场中的数据共享、交换和流通奠定坚实的数据基础。

首先，卫生健康数据通过节点模块进入数据连接模块，基于数据件封装技术，数据要素形成标准数据件。其次，数据交换模型将不同数据节点的数据件按照数据标准体系进行转换，将数据件映射至相同话语体系。再次，在数据管理平台上，利用跨域数据管理技术，将多源异构的数据件转化成语义统一的融合数据，同时穿透式安全技术对数据操作进行行为存证、对数据进行可信溯源，实现数据使

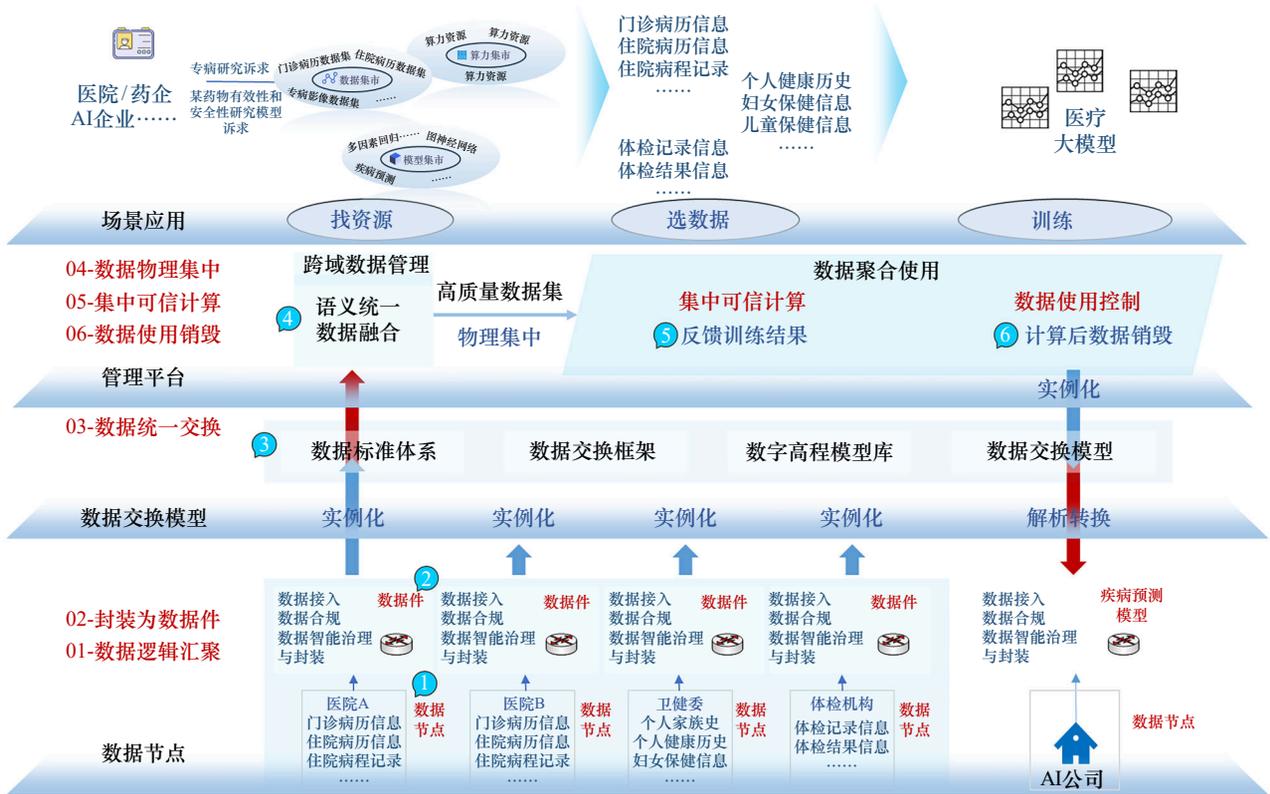


图5 卫生健康数据要素场工程实践
注：AI为人工智能，卫健委为卫生健康委员会。

用的全链路管控，形成高质量数据集。对于聚合后的数据，管理平台结合具体医疗应用场景（如训练医疗大模型），一方面结合供需关系，挑选、推荐符合适于构建医疗大模型的数据，另一方面借鉴聚变式处理技术与低熵化流通技术，为不同数量、不同类别的医疗数据明晰价值，完成卫生健康数据典型场景的应用。最后，为了保障数据安全的出域和价值转化，数据要素场可以按需将计算后的医疗数据进行销毁。

卫生健康数据在数据要素场内完成流通，基于利用高质量数据集训练的医疗大模型计算的结果，在可用性和准确性上更加符合医疗、体检、医药研发、科学研究等卫生健康领域的应用需求，真正做到数据要素流通的可落地，形成可借鉴、可复用的标准化工程实践案例。

在工程实践中，除了数据的连接、流通和使用以外，数据服务与价值释放是更加重要的落地实践重点。数据的运营服务和数据价值释放需要引入复杂计算、大模型和人工智能算法，对海量数据进行挖掘和分析，生成有价值的洞察和预测结果，这些

结果可以用于智能决策支持、业务流程优化、精准营销等实际应用场景，构建起数据驱动的智能决策支持系统，辅助各领域的决策者作出科学、精准的判断。

在当前的数据要素应用环境中，面临着诸如数据质量和标准化不足、“数据孤岛”现象严重以及数据安全与隐私保护挑战等问题，针对这些问题，卫生健康数据要素场工程实践通过系统化的规划和分阶段实施，通过数据标准化高质量汇集、数据要素高效流通、数据要素高价值应用等方向的努力，解决了数据要素流通利用领域面临的实际问题。

五、数据要素场未来发展展望

随着数据要素场理论和技术体系的不断发展，数据要素场落地应用的不断深入、案例不断增加，未来的研究和实践将在基础理论、关键技术、工程实践和生态建设等方面取得突破，并进一步推动数字经济的发展和社会治理的现代化。

（一）深入研究数据要素场基础理论

数据要素场的发展离不开底层基础理论的指导。未来，数据要素在社会生产和经济活动中的重要性日益突出，数据要素场底层驱动机理、要素协同内涵和价值释放模式的明晰与理解成为关键。政府和行业先驱者将进一步完善数据场基础理论，包括数据场的形成机理、数据要素间的价值作用关系、数据价值的释放模式等，厘清数据场底层基础理论在实践中的指导作用，为数据要素场技术体系建设和工程实践应用提供科学依据。

（二）实现数据要素场关键技术突破创新

关键技术创新将继续驱动数据要素场的发展。数据要素场关键技术应致力于构建和优化数据场，解决数据场中数据流通和价值释放等难题。未来，隐私计算、联邦学习、区块链等先进技术将进一步优化、提升数据流通和共享的效率与安全性，包括加速发展数据采集与融合技术、数据存储与管理技术、数据高效流通技术、数据安全保障技术等。数据要素场关键技术的创新与发展将推动各行业的数据驱动创新和数字化转型，为未来实现智能社会奠定坚实基础。

（三）推动数据要素场工程实践应用

工程实践应用是检验数据要素场基础理论和技术体系的有效标准。数据场在经济领域、社会治理、公共服务等方面发挥重要价值与应用意义。未来，数据场将为智慧城市、智能制造、智慧农业等提供有力支持，更多的行业和领域在工程实践中通过数据场技术实现数字化转型，提升生产效率和服务水平。基于数据要素场基础理论指导和技术体系支撑下工程实践的推广应用，社会各界将共同受益，进一步实现数据的社会价值和经济效益。

（四）建设数据要素场生态

稳定数据生态建设是数据要素场发展的核心目标。数据要素场的生态建设将致力于构建一个开放、协同、共享的生态系统，通过完善的数据基础理论、先进的技术治理体系和典型的工程实践应用，促进数据要素的自由流通和高效利用。未来，跨行业、跨领域的合作将得到加强，数据资源得到充分整合与挖掘，推动各行业数字化转型与创新发

展，形成一个兼具制度规范性与技术先进性的数据要素生态体系，实现数字社会的美好愿景。

利益冲突声明

本文作者在此声明彼此之间不存在任何利益冲突或财务冲突。

Received date: November 2, 2024; **Revised date:** December 2, 2024

Corresponding author: Hong Richang is a professor from School of Computer Science and Information Engineering, Hefei University of Technology. His major research field is artificial intelligence. E-mail: hongrc@hfut.edu.cn

Funding project: Chinese Academy of Engineering project “Research on the Development Strategy of National Data Space” (2024-XBZD-05)

参考文献

- [1] 中共中央 国务院. 国务院关于印发“十四五”数字经济发展规划的通知 [EB/OL]. (2021-12-12)[2024-10-28]. https://www.gov.cn/zhengce/content/2022-01/12/content_5667817.htm.
The Central Committee of the Communist Party of China and the State Council. Notice of the State Council on issuing the “14th Five Year Plan” for the development of the digital economy [EB/OL]. (2021-12-12)[2024-10-28]. https://www.gov.cn/zhengce/content/2022-01/12/content_5667817.htm.
- [2] 中共中央 国务院. 关于构建数据基础制度更好发挥数据要素作用的意见 [EB/OL]. (2022-12-02)[2024-10-28]. https://www.gov.cn/zhengce/2022-12/19/content_5732695.htm.
The Central Committee of the Communist Party of China and the State Council. Opinions on building a data infrastructure system to better utilize the role of data elements [EB/OL]. (2022-12-02)[2024-10-28]. https://www.gov.cn/zhengce/2022-12/19/content_5732695.htm.
- [3] 中共中央办公厅 国务院办公厅. 关于加快公共数据资源开发利用的意见 [EB/OL]. (2024-09-21)[2024-10-28]. https://www.gov.cn/zhengce/202410/content_6978911.htm.
The General Office of the Central Committee of the Communist Party of China and the General Office of the State Council. Opinions on accelerating the development and utilization of public data resources [EB/OL]. (2024-09-21)[2024-10-28]. https://www.gov.cn/zhengce/202410/content_6978911.htm.
- [4] 国家发展与改革委员会. 关于完善数据流通安全治理 更好促进数据要素市场化价值化的实施方案 [EB/OL]. (2025-01-06)[2025-01-15]. https://www.ndrc.gov.cn/xxgk/zcfb/tz/202501/t20250115_1395692.html.
National Development and Reform Commission and other department. Implementation plan for improving data circulation and security governance to better promote the marketization and valorization of data elements [EB/OL]. (2025-01-06)[2025-01-15]. https://www.ndrc.gov.cn/xxgk/zcfb/tz/202501/t20250115_1395692.html.
- [5] Congress.gov. Text-H.R. 4174-115th congress (2017—2018): An act to amend titles 5 and 44, United States Code, to require Federal evaluation activities, improve Federal data management, and for other purposes [EB/OL]. (2019-01-14)[2024-12-02]. <https://www>

- congress.gov/bill/115th-congress/house-bill/4174/text.
- [6] U. S. Office of Management and Budget. Federal data strategy 2021 action plan [EB/OL]. (2021)[2024-12-02]. <https://strategy.data.gov/assets/docs/2021-Federal-Data-Strategy-Action-Plan.pdf>.
- [7] European Commission. 2030 digital compass: The European way for the digital decade [EB/OL]. (2021-03-09)[2024-12-02]. <https://eufordigital.eu/wp-content/uploads/2021/03/2030-Digital-Compass-the-European-way-for-the-Digital-Decade.pdf>.
- [8] Department for Digital, Culture, Media & Sport. UK digital strategy [EB/OL]. (2022-10-04)[2024-12-02]. <https://www.gov.uk/government/publications/uks-digital-strategy/uk-digital-strategy>.
- [9] Government of Ireland. The digital ireland framework [EB/OL]. (2022-02-01)[2024-12-02]. <https://www.gov.ie/pdf/?file=https://assets.gov.ie/214584/fa3161da-aa9d-4b11-b160-9cac3a6f6148.pdf#page=null>.
- [10] 杜小勇. 对数据要素的几点认识 [J]. 农业大数据学报, 2023, 5(1): 8–10.
Du X Y. Preliminary exploration of data factors [J]. *Journal of Agricultural Big Data*, 2023, 5(1): 8–10.
- [11] 刘博文, 夏义堃. 基于数据空间的产业数据流通利用: 逻辑框架与技术实现 [J]. 图书与情报, 2024 (2): 33–44.
Liu B W, Xia Y K. Industrial data circulation and utilization based on data spaces: Logical framework and technical implementation [J]. *Library & Information*, 2024 (2): 33–44.
- [12] 杜小勇, 李彤, 卢卫, 等. 跨域数据管理 [J]. 计算机科学, 2024, 51(1): 4–12.
Du X Y, Li T, Lu W, et al. Cross-domain data management [J]. *Computer Science*, 2024, 51(1): 4–12.
- [13] Sheath A P, Larson J A. Federated database systems for managing distributed, heterogeneous, and autonomous databases [J]. *ACM Computing Surveys*, 1990, 22(3): 183–236.
- [14] Guo J Q, Zhan Z C, Gao Y, et al. Towards complex text-to-SQL in cross-domain database with intermediate representation [R]. Florence: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019.
- [15] Zheng Y. Methodologies for cross-domain data fusion: An overview [J]. *IEEE Transactions on Big Data*, 2015, 1(1): 16–34.
- [16] Xiao Z, He D, Du J Y. A Stackelberg game pricing through balancing trilateral profits in big data market [J]. *IEEE Internet of Things Journal*, 2021, 8(16): 12658–12668.
- [17] Chen Z H, Wang C, Wang Q, et al. Dynamic budget throttling in repeated second-price auctions [R]. Palo Alto: The AAAI Conference on Artificial Intelligence, 2024.
- [18] Duan Z J, Tang J W, Yin Y T, et al. A context-integrated transformer-based neural network for auction design [R]. Baltimore: International Conference on Machine Learning, 2022.
- [19] 程啸. 区块链技术视野下的数据权属问题 [J]. 现代法学, 2020, 42(2): 121–132.
Cheng X. On ownership of data in the blockchain [J]. *Modern Law Science*, 2020, 42(2): 121–132.
- [20] 江小涓, 宫建霞, 李秋甫. 数据、数据关系与数字时代的创新范式 [J]. 中国社会科学, 2024 (9): 185–203.
Jiang X J, Gong J X, Li Q F. Data, data relations and innovation paradigms in the Digital age [J]. *Social Sciences in China*, 2024 (9): 185–203.
- [21] Adi Y, Baum C, Cisse M, et al. Turning your weakness into a strength: Watermarking deep neural networks by backdooring [R]. Berkeley: 27th USENIX security symposium (USENIX Security 18), 2018.
- [22] Mohassel P, Secureml Z Y. A system for scalable privacy-preserving machine learning [R]. Los Alamitos: 2017 IEEE Symposium on Security and Privacy (SP), 2017.
- [23] Demmler D, Schneider T, Zohner M. ABY-A framework for efficient mixed-protocol secure two-party computation [R]. Reston: The 2015 Network and Distributed System Security, 2015.
- [24] Carlini N, Wagner D. Towards evaluating the robustness of neural networks [R]. Los Alamitos: 2017 IEEE Symposium on Security and Privacy, 2017.