人工智能背景下数据治理体系研究

陈春梅
拉萨龙科电子科技有限公司　西藏拉萨　850000

摘　要：人工智能的快速发展对传统数据治理体系提出了全新挑战，本研究探讨了AI背景下数据治理的内涵演变，分析了技术驱动的治理难题，包括数据规模复杂性、算法黑箱问题和实时性要求。研究构建了包含主权确认、全生命周期管理、算法透明化等要素的新型治理框架，并提出智能化技术体系、算法治理、跨境协作等优化路径。通过研究，论证了技术与制度协同的必要性，为构建安全、高效、可信的AI数据治理体系提供理论支撑和实践指导。

关键词：人工智能；数据；治理体系

引言

随着人工智能技术深度应用，数据已成为核心生产要素，但数据滥用、隐私泄露等问题日益凸显。传统治理模式难以应对AI场景下的动态性、复杂性挑战。本文基于数据主权和算法伦理理论，系统研究AI与数据治理的互动关系，重点探讨数据价值挖掘与风险防控的平衡机制，以及适应AI特性的治理框架构建路径。通过分析技术与制度的协同关系，提出了兼顾创新发展与安全可控的治理方案。

一、数据治理的内涵

数据治理是指通过建立系统的管理框架，确保数据在其全生命周期中的可用性、完整性、安全性和合规性。这一内涵强调数据不仅仅是静态的信息资产，而是需要动态管理的战略资源。系统性管理框架包括数据标准制定、质量控制、权限分配和流程优化等多个维度。例如，在金融行业，数据治理要求统一客户数据的定义和存储格式，避免因数据冗余或错误导致决策偏差。同时，数据治理需要跨部门协作，确保业务、技术和合规团队在数据使用上达成一致。其核心目标是提升数据价值，降低风险，并支持企业战略目标的实现。

数据治理的另一重要内涵是平衡数据利用的权责关系与伦理约束，随着人工智能和大数据技术的普及，数据主体（如个人用户）的隐私权、数据控制者的使用权以及监管机构的监督权之间可能产生冲突。数据治理需明确各方责任，例如企业需履行数据保护义务，政府需制定合理的监管规则，而用户应具备知情权和选择权。此外，伦理问题如算法歧视、数据垄断等也需纳入治理范畴。因此，数据治理不仅是技术问题，更是涉及法律、伦理和社会责任的综合性体系。

二、技术驱动的治理挑战

1.数据规模与复杂性带来的治理难题

人工智能的广泛应用导致数据规模呈指数级增长，数据类型也从传统的结构化数据扩展至文本、图像、视频等非结构化数据。海量数据的存储、处理和分析对传统数据治理体系提出了严峻挑战。例如，在自动驾驶领域，车辆每天产生的传感器数据可达数TB，如何高效清洗、标注并存储这些数据成为关键问题。同时，多模态数据的融合分析增加了数据关联的复杂性，可能引发数据冗余、不一致甚至错误传播的风险。数据治理必须适应这种规模化和复杂化的趋势，否则将难以保障数据质量，进而影响AI模型的训练效果和决策可靠性。

2.算法黑箱与数据可解释性不足

人工智能模型，尤其是深度学习算法，通常以“黑箱”方式运行，导致数据输入与输出之间的逻辑关系难以追溯。这种不可解释性给数据治理带来了新的挑战。例如，在金融风控领域，AI模型可能因数据偏差而拒绝某些群体的贷款申请，但由于算法决策过程不透明，监管机构难以验证其公平性。此外，数据在模型训练过程中的流转路径也难以监控，可能掩盖数据滥用或隐私泄露的风险。数据治理需要建立算法审计机制，确保数据使用过程可追溯、可解释，从而降低因算法不透明带来的合规与伦理风险。

3.实时数据处理与动态治理需求

人工智能的许多应用场景依赖实时数据流，如智能城市中的交通监控或工业物联网中的设备预测性维护。传统的数据治理框架通常针对静态数据设计，难以适应实时数据的动态特性。例如，在实时推荐系统中，用户

行为数据被持续采集并即时反馈至模型，数据治理必须同步进行质量检测和异常监控，否则错误数据可能迅速影响推荐结果。此外，实时数据的跨境流动也增加了合规难度，不同司法管辖区的数据监管要求可能随时变化。数据治理体系需具备动态调整能力，确保在高速数据流动中仍能维持安全性、准确性和合规性。

三、人工智能背景下的数据治理框架构建

1.数据主权与多方协同治理机制

在人工智能时代，数据治理框架的首要任务是明确数据主权归属并构建多方协同治理机制。数据主权不仅涉及国家层面的数据跨境流动管制，还包括企业、个人对数据的控制权。例如，在医疗健康领域，患者的基因组数据既涉及个人隐私，又具有科研价值，需要明确患者、医疗机构、研究机构之间的权责划分。协同治理机制要求政府、企业、行业组织及公众共同参与，通过数据治理委员会或伦理审查机构等形式，制定兼顾效率与公平的规则。欧盟《数据治理法案》提出的“数据中介服务”便是一种尝试，旨在通过第三方机构促进数据共享的同时保护各方权益。这一机制的核心在于平衡数据利用的公共利益与个体权利，避免因权力集中导致的数据垄断或滥用。

2.全生命周期数据质量管理体系

人工智能模型高度依赖数据质量，因此数据治理框架必须覆盖数据的全生命周期，从采集、存储、处理到应用和销毁。在数据采集阶段，需建立标准化规范，确保数据来源合法且具有代表性，例如自动驾驶领域需明确传感器数据的采集频率和精度要求。在数据处理阶段，需引入自动化清洗和标注工具，减少人为错误，同时通过元数据管理追踪数据变更历史。数据应用阶段则需持续监控模型输出与数据质量的关联性，例如金融风控系统需定期检测数据偏差对算法公平性的影响。最后，数据销毁阶段需符合隐私保护要求，如匿名化或加密删除。全生命周期管理的目标是构建闭环治理体系，确保数据在每个环节均满足可用性、一致性与合规性要求，从而提升AI系统的可靠性与可信度。

3.算法透明化与责任追溯技术

人工智能的“黑箱”特性是数据治理的核心挑战之一，因此框架需整合算法透明化技术与责任追溯机制。算法透明化要求模型开发者提供可解释性报告，例如通过特征重要性分析或决策路径可视化，使监管方和用户理解数据如何影响输出。在司法领域，AI辅助量刑系统必须公开其依据的法律条文和案例数据，避免因算法不透明引发公正性质疑。责任追溯则依赖区块链、数字水印等技术，记录数据使用链条与算法决策过程。例如，在版权保护中，区块链可追溯训练数据的来源，防止未经授权的数据被用于商业模型。此外，需建立算法备案制度，要求高风险AI系统提交数据使用说明和伦理评估，确保事后追责有据可依。

4.隐私计算与数据安全技术嵌入

数据治理框架需深度集成隐私计算与安全技术，以解决人工智能场景下的隐私泄露与数据滥用风险。隐私计算技术如联邦学习、多方安全计算和差分隐私，能够在数据不直接共享的前提下实现联合建模与分析。例如，医疗领域的跨机构研究可通过联邦学习利用各医院的病例数据，而无需集中上传原始数据。数据安全技术则包括同态加密、访问控制和入侵检测系统，确保数据在存储、传输和处理中的保密性与完整性。在智慧城市项目中，居民行为数据需经加密后供AI模型分析，同时严格限制政府与企业的访问权限。技术嵌入的关键在于与治理规则联动，例如根据数据敏感等级动态调整加密强度或计算方式。这种技术驱动的治理模式，能够在保障数据安全的同时释放其价值。

5.动态合规与跨境数据流动规则

人工智能的全球化应用要求数据治理框架具备动态适应性，尤其是应对跨境数据流动的合规需求。不同司法管辖区对数据主权、隐私保护和本地化存储的要求差异显著，例如欧盟GDPR强调“数据最小化”，而中国《数据安全法》要求关键数据境内存储。治理框架需设计弹性规则，例如通过“数据沙盒”允许企业在可控环境中测试跨境数据方案，或建立白名单机制简化低风险数据流动的审批流程。同时，需推动国际标准协同，如APEC跨境隐私规则（CBPR）与欧盟充分性认定的互认。在具体场景中，自动驾驶汽车的跨国行驶可能需实时传输路况数据，治理框架应明确数据分级（如非敏感数据可跨境，高精地图数据需本地处理）和应急响应机制。动态合规的核心是通过技术工具（如自动化合规检测平台）与制度创新，降低企业跨境数据业务的合规成本。

四、优化路径

1.构建智能化数据治理技术体系

人工智能技术的快速发展为数据治理提供了全新的技术手段，构建智能化数据治理技术体系成为优化路径的首要任务。这一体系需要整合机器学习、自然语言处理、知识图谱等AI技术，实现数据治理的自动化与智能化。例如，通过机器学习算法可以自动识别数据中的异

常值和缺失值，大幅提升数据清洗效率；利用自然语言处理技术能够对非结构化文本数据进行智能分类和标注，解决传统人工处理效率低下的问题。知识图谱技术则可以将分散在不同系统中的数据关联起来，形成统一的数据资产视图，为数据治理决策提供支持。在具体实施层面，企业需要建设智能数据治理平台，集成数据质量检测、元数据管理、数据血缘分析等功能模块，并通过AI模型持续优化治理规则。政府部门也应推动建设国家级数据治理基础设施，如公共数据开放平台和行业数据共享中心，为智能化治理提供基础支撑。

2.完善数据确权与利益分配机制

明确数据权属和建立公平合理的利益分配机制是优化数据治理的关键路径，当前数据要素市场面临的核心困境在于数据产权界定不清，导致数据流通受阻和价值创造受限。完善数据确权机制需要从法律层面明确数据所有权、使用权、收益权的归属原则，区分个人数据、企业数据、公共数据等不同类型数据的权利边界。例如，对于用户行为数据，应当承认平台企业的数据使用权，同时保障用户对个人数据的控制权和收益权。在利益分配方面，可探索建立数据要素参与分配的具体实现形式，如数据分红、数据股权等创新模式。区块链技术的应用能够为数据确权和利益分配提供技术支持，通过智能合约自动执行数据交易和收益分配。

3.强化算法治理与伦理审查制度

随着算法在社会各领域的深度应用，强化算法治理成为优化数据治理体系的重要路径。这需要建立覆盖算法设计、开发、部署、应用全过程的治理机制，重点解决算法歧视、算法垄断等新型治理难题。在制度设计上，应当建立算法分级分类管理制度，对医疗诊断、金融信贷等高风险算法实施备案审查，对推荐系统等低风险算法实行事后监管。算法透明性要求应当具体化，包括公开算法基本原理、训练数据特征、性能指标等信息，接受社会监督。同时，需要设立专门的算法伦理审查委员会，由技术专家、法律专家、伦理学者等共同参与算法伦理风险评估。

4.培育数据素养与公众参与机制

提升全社会数据素养和健全公众参与机制是优化数据治理的基础性工程，当前数据治理面临的重要挑战是公众认知不足和参与渠道缺失，导致治理决策缺乏广泛代表性。培育数据素养需要将数据教育纳入国民教育体系，在中小学开设数据科学基础课程，在高校加强数据伦理和法律教育。面向社会公众开展数据知识普及活动，提高对数据权利和风险的认知水平。在公众参与机制建设方面，可以借鉴环境治理中的公众参与经验，建立数据治理听证会、民意调查、社区议事等制度。例如，在城市智慧化建设中，应当就人脸识别等技术的应用广泛征求市民意见。数字平台应当为用户提供便捷的数据控制工具，如个性化隐私设置、数据使用授权管理等。同时，支持消费者组织、行业联盟等第三方机构参与数据治理监督。培育数据素养和公众参与机制能够增强数据治理的民主性和正当性，形成政府监管、企业自律、公众监督的多元共治格局。

结束语

人工智能背景下的数据治理需要技术革新与制度设计的深度融合，本研究揭示了AI对治理体系的颠覆性影响，提出了动态化、协同化的解决方案。未来研究可进一步探索量子计算等新技术带来的治理范式变革，以及数据产权制度的细化路径。只有持续完善治理体系，才能确保人工智能在合规框架下释放最大社会价值，推动数字经济高质量发展。

参考文献

[1]李森.风险防范视阈下生成式人工智能数据安全的治理路径——以GPT类模型为例[J].西藏民族大学学报(哲学社会科学版)，2023，44(06)：139-145.

[2]针晓东.论生成式人工智能的数据安全风险及回应型治理[J].东方法学，2023，(05)：106-116.

[3]张楠.人工智能提升数据治理智能化水平[J].软件和集成电路，2023，(08)：30.

[4]张浩.人工智能治理的实践进展与展望[J].人工智能，2022，(01)：16-21.

[5]张臻.智能时代的教育数据治理变革：挑战与路径[J].中国教育信息化，2022，28(01)：11-17.

[6]张述存.“十四五”时期推进社会治理智能化的几点思考[J].社会治理，2022，(01)：46-48.

[7]韦苇，任锦鸾，杨青峰.短视频平台数据治理框架和机制研究[J].电子政务，2022，(04)：64-72.