

基于区块链的人工智能大数据交易可信溯源体系构建

陈斌

杭州国际数字交易中心有限公司, 浙江省杭州市, 310051;

摘要: 随着人工智能技术的快速发展, 大数据资源作为其基础驱动力, 正在成为关键性战略资产。然而, 当前大数据交易市场中普遍存在数据来源不清、交易过程不透明、使用权限无法溯源等问题, 严重制约了数据要素市场的高效流通与可信交互。区块链技术以其去中心化、不可篡改、全流程可追溯等技术特性, 为构建安全透明的大数据交易环境提供了可能。本文以人工智能大数据交易为核心场景, 系统分析现有数据流通体系存在的信任缺失问题, 探索基于区块链的可信溯源模型构建路径, 重点阐述数据确权、链上登记、智能合约管控与使用行为溯源等关键机制, 并提出构建“数据可信+使用可控+收益可证”的新型数据交易架构。研究认为, 将区块链与AI数据管理深度融合, 不仅能重构数据要素交易信任机制, 更是推动数据资产化、市场化和可持续流通的有效手段。

关键词: 区块链; 人工智能; 大数据交易; 可信溯源; 智能合约; 数据确权

DOI: 10.64216/3080-1508.25.03.039

引言

在人工智能与大数据深度融合的背景下, 数据正逐步成为与资本、土地并列的核心生产要素, 其流通与交易能力已成为衡量数字经济发展的标志。然而, 当前大数据交易在实践中仍面临信任断层、信息不对称、确权困难等诸多痛点。例如, 数据提供方难以证明数据来源合法, 数据使用方无法明确使用权限边界, 监管方则难以追溯异常行为或违规流转, 严重制约了数据要素市场的规范化发展。与此同时, 数据一旦泄露或滥用, 往往难以追责, 进一步加剧市场参与者之间的信任赤字。

区块链技术的出现为解决上述问题提供了新的技术范式。其核心特性——去中心化、分布式账本、智能合约和可追溯性等, 为建立数据可信确权、透明交易、行为可审计的体系提供了可靠支撑。尤其在人工智能领域, 模型训练对数据质量、来源及合法性要求极高, 因此构建一个具备“可审计、可控、可验证”能力的大数据交易溯源体系, 是释放AI潜力、激活数据价值的关键。本文从问题现状出发, 提出一种基于区块链技术的AI大数据可信溯源交易框架, 通过技术创新构建多方可信、全链留痕的数据生态环境, 为数据要素市场发展提供可落地路径。

1 大数据交易中的信任缺失与监管困境

1.1 数据来源不明导致交易信用薄弱

在当前大数据交易实践中, 数据提供者往往无法提供明确的来源证明, 导致数据的真实性、完整性难以验证。尤其是面对海量异构数据, 如社交媒体数据、用户

行为日志、医疗健康数据等, 其采集路径复杂, 归属权模糊, 数据授权使用的边界模糊。一旦在人工智能模型训练中引入来源不明或侵权数据, 不仅可能导致训练结果失真, 更可能引发重大合规和伦理问题。此外, 部分中间平台在未授权的情况下转卖数据, 更加剧了整个市场对“黑数据”流通的担忧, 极大损害了数据交易市场的信誉基础。

1.2 数据交易链条长、参与方信息不对称

大数据交易链条通常涉及数据提供者、集成服务商、应用开发商、监管机构等多个参与方, 每一环节若缺乏有效信任机制和信息共享标准, 极易形成“黑箱操作”与资源错配。当前多数数据交易仍依赖于传统平台撮合或人工签约方式, 合同执行效率低下、违约成本高、权责边界模糊。而对于使用方而言, 数据质量、更新频率、敏感字段处理等关键信息常常无法全面掌握, 导致使用效果与预期严重偏差。此外, 数据交易过程中若缺乏全链条记录机制, 数据使用行为难以审计, 既无法保障数据所有者权益, 也无法为监管提供有效证据。

1.3 缺乏统一标准导致监管难以落地

尽管近年来政府和行业机构纷纷出台大数据交易指导规范, 但由于缺乏跨平台、跨行业的数据登记与验证机制, 现有标准多呈“碎片化”状态, 难以有效覆盖整个数据生命周期管理过程。一方面, 数据确权、脱敏处理、交易定价等核心环节缺乏技术工具支撑, 造成实践中落地困难; 另一方面, 监管部门由于缺乏统一的数据链路视图, 往往只能事后干预, 难以进行实时监管与

动态审计。这种“监管缺位”的状态，进一步助长了灰色交易和违法行为的滋生，使数据交易在迈向规模化发展的过程中始终难以实现高质量合规运营。

2 区块链技术赋能数据可信溯源机制重塑

2.1 链上数据确权保障资产归属可验证

数据确权是可信数据交易体系必须的步骤，但传统方法主要依靠纸质合同或平台签合同，或者行政认定的方式，会有成本高、效率不好和容易被修改这些问题。区块链因为有时间戳，还有分布式的账本结构和不能被修改的属性，所以它给数据确权带来了更可靠的办法。具体来说，数据提供方可以把原始数据做哈希处理然后传到区块链上，生成唯一的数据摘要当数字指纹。同时把数据的元信息比如采集时间、设备、使用场景这些都一起固定在区块链里，这样就能确定数据资产最开始是谁的，还有它的合法性。

在区块链上这些信息可以被每个节点查看到，保证确权过程是公开透明还不能随便否认的。比如说医院收集电子病历时，可以马上把数据加密放进区块链，再给每个数据发专门的 ID 号，让数据和数字身份完全对应起来。这样做不仅能防止数据被伪造复制，还可以在数据交易或分享时给出法律依据，让使用数据的人更相信数据的质量和合规性。另外区块链确权还能和司法上的证据保存系统合作，以后要是出现纠纷就可以用区块链上的记录当证据，真正让数据归属可以看得见、能验证、能证明。

2.2 智能合约实现权限控制与行为合规

在传统的数据交易里，合同的执行必须要依靠人工监督和中间平台，这样不仅效率很低容易违约，还会在使用时发生权限乱用或者信息泄露的问题。区块链里的智能合约能把交易规则写成代码自动执行，让整个数据交易过程变得更自动化和可以控制。在人工智能大数据交易的时候，智能合约可以根据不同数据资产的使用类型和使用者的身份，设置多个方面参数，比如授权有多大范围能用多少次时间多久怎么调用还有是否可以二次分发，同时把付款方式和违规处理办法绑定起来，如果有操作超过了规定，立马就会停止访问数据，然后在区块链上记录下来，确保使用行为不会超出合同约定。而且智能合约还能帮买卖双方直接结算，比如说根据调用次数算钱马上打到数据提供者的账户里，完全不用像以前人工对账那样耽误时间还容易出错。这种机制还能支持按使用次数付费、分阶段开放数据、授权就扣钱这些灵活的买卖方式，不但提高了交易速度还降低了信任

成本，特别适合那些规模很大、交易频繁、参与方很多的人工智能训练场景。更重要的是，智能合约作为区块链程序的一部分，它的执行记录都能查到，有天生的可以被审查的特点，给整个合约执行过程提供了公开透明的信任证明。

2.3 全链溯源机制强化交易行为可审计

在现在的数据交易系统里，缺少完整记录数据流转整个过程和统一审计的办法，这样当发生数据泄露或者模型出问题，很难找到数据来源和该谁负责。区块链的追溯功能正好能解决这个问题，它把所有数据交易动作都记在区块链上，实现从数据上传、使用、传输、处理到输出结果的全流程记录。在 AI 使用数据时，每次查看数据都会像交易记录一样写在区块链上，包括谁查看了、什么时候查的、用了什么接口、用来做什么这些信息，这些记录都不能随便改动或者删除，保证交易路径是真实完整的。

如果有外界对 AI 模型结果提出合规或伦理质疑，开发者或监管方就能通过区块链上的日志快速弄清楚整个数据是怎么被调用的，看看有没有越权操作、数据滥用或者脱敏处理没做好这些情况。为了让不懂技术的监管人员也能看懂，系统可以配上图形化的查看工具搞个‘数据追踪地图’，把数据在不同阶段的流转路线和参与方都显示出来，这样监管起来更方便更有效率。以后要是区块链记录数据的方式变得更标准化、更直观的话，就能帮助建立既不容易被篡改、又能追溯行为、还能明确责任的可信数据交易审查体系。

3 人工智能大数据可信交易体系构建路径

3.1 构建“可信数据空间”实现链上数据资产化

构建‘可信数据空间’是实现人工智能数据资源变成价值和可控使用的基础的核心部分。这不只是个技术上的东西，同时还有数据管理、合规监督和经济奖励这些方面结合起来的一个体系。在这里框架里，所有原始数据要放进来的话都得先做区块链登记和确定权利，用区块链给每个数据打上唯一的链上记号，还要和数据提供者的数字身份紧紧绑在一起，这样就能确定数据是谁的、从哪来的没问题。通过在元数据层里写上数据的重要特点，比如有哪些敏感字段、可以用在什么地方、能用多久、服务等级怎么样这些信息，可信数据空间就能给 AI 模型提供有结构、有标准而且透明的数据资源。

同时，为了解决区块链存储性能和隐私保护的问题，该系统用了“链上登记、链下保存”的方法，数据保存在分布数据库或云端加密平台，只把数据摘要和权限信

息放到链上,这样既确保数据完整性没有问题,同时不会增加存储压力。另外系统还加入了智能合约的模板、审计接口调用和实时结算这些功能,覆盖数据从接入授权到调用赚钱的整个流程,做到数据可以共享但看不见内容。作为人工智能公司和开发者的基础数据服务平台,可信数据空间不光保证了数据合规和安全,还让模型训练的数据质量更好更容易解释,给人工智能发展带来更强更可信的基础支持。

3.2 推动数据交易联盟链生态共建

现在数据交易市场有很多问题,比如平台都是分开的、规则都不一样、还有各自利益保护,这些导致数据流动不顺畅,重复建设太多,交易时候信任成本太高了。搞一个数据交易联盟链生态的话,就能解决这些限制,让数据流动更高效还有管理更可信。联盟链作为大家一起建设的分布式记账系统,应该在顶层设计的时候把技术、制度和治理规则这些方面都协调好,具体来说应该让政府部门管数据合规和安全的事,行业里的龙头企业来提供平台和应用场景,大学和研究机构搞标准制定和协议研究,技术公司做底层技术和接口开发,审计单位专门监督链上交易。这种联盟链用的是“多中心一起治理”模式,既让各个节点自己管理数据存储,又通过跨链通信建了好多数据共享通道,这样就能跨行业跨地区跨系统做交易了。

技术上来说,联盟链要有标准化的智能合约接口,这样数据调用协议就能快速部署和重复使用,还要加入信用评分、黑名单、数据评级这些工具,鼓励提供好数据,打击恶意行为。生态建设方面,要让中小企业和开发者都能参与进来,别让数据都被大公司垄断,形成“规则严格+参与广泛+收益多样”的合作型数据交易生态。这样的话,联盟链就能建立既能保护数据产权、规范交易流程,又能激活市场促进数据创新的数字经济基础了。

3.3 构建以用户为中心的数据权益保护机制

数据作为“新的生产原料”,它们归属权和使用权分开的情况越来越明显,怎么能让数据流动起来同时又保证提供数据的人权益不被损害,这变成可信的交易系统需要解决的主要问题。建立以用户为核心的数据权益保护机制,关键在于要让每个数据主体都对他们的数据有清晰的控制权和可以感知的使用反馈权利。在区块链支持下,每个用户都能通过数字身份认证得到一个唯一的链上身份,并以此对自己的数据做精细化的授权设置。系统可以让用户按照数据的内容、使用场景、访问者的

类型等多个方面来设置授权规则,比如限制有些健康数据只能用于非商业的医疗研究,并且还可以设置有效时间和最多被调用的次数。在使用的时候,用户能从系统后台随时查看他们的数据使用记录,里面有调用时间、调用是谁、用来做什么这些重要信息,从而让整个数据使用过程可以看得见。

同时智能合约让用户在数据交易中能够按照合约规定比例得到使用收益,比如根据数据调用次数、使用频率还有敏感程度这些指标来设定收益模型,并且将交易的分成自动打回用户的账户里。另外系统还允许用户可以随时撤销他们的授权或者更新授权内容,这样就能防止那种一旦授权就永远不能改可能被滥用的风险。通过这些设计,用户不再是只能被动提供数据的人相反他们变成了有权知道信息情况自己决定事情还能参与经济利益分配的主体。这种以人为本的设计思路不仅仅能让用户更信任整个系统同时还能帮AI系统的训练创造一个更健康公平能长久持续的数据供应环境,这样一来数据生态就能长期繁荣发展下去。

4 结语

面对数据要素价值化与交易市场快速发展所带来的信任瓶颈问题,构建以区块链为基础的人工智能大数据可信溯源体系已成为数字经济时代的重要课题。本文系统梳理了当前数据交易中的核心挑战,并提出以数据确权、智能合约、全链追踪为支撑的溯源机制设计思路,为数据资产在流通过程中的安全、可信与可控提供了实践路径。未来,随着区块链技术与AI算法的不断融合,将有望进一步推动数据资产确权标准、数据价值评估机制以及跨平台信任协同系统的成熟落地,真正实现从“可交易”到“可信任”的数据经济跃升。

参考文献

- [1] 李晓黎. Solidity 智能合约开发技术与实战[M]. 人民邮电出版社:202212. 302.
- [2] 丁晓蔚. 从互联网金融到数字金融: 发展态势、特征与理念[J]. 南京大学学报(哲学·人文科学·社会科学), 2021, 58(06): 28-44+162.
- [3] 刘光强. 基于区块链审计的智能审计研究[J]. 商业会计, 2021, (13): 4-14.
- [4] 朱晓武, 黄绍进. 数据权益资产化与监管[M]. 人民邮电出版社:202009. 189.
- [5] 陈珊珊, 钟燕. 人工智能在证券交易系统应用中面临的问题及对策[J]. 现代信息科技, 2019, 3(07): 134-136.