基于区块链的AI模型数据安全共享方案研究

Research on Data Security Sharing Scheme for AI Model Based on Blockchain

程筱彪，张曼君，谢泽铖（中国联通研究院，北京 100048）
Cheng Xiaobiao，Zhang Manjun，Xie Zecheng（China Unicom Research Institute，Beijing 100048，China）

摘　要：
AI模型的训练和优化高度依赖海量数据，这些数据的来源呈现显著的多样性与分散性，传统的数据共享机制面临着数据泄露、数据篡改、隐私保护等问题。在此背景下，构建AI模型数据安全共享方案成为当务之急。该方案依托区块链与同态加密技术，可确保训练数据无需离开本地，通过联盟协作训练实现AI模型开发。通过这种方式，能在更大范围内实现数据的安全共享，进而充分释放AI模型的巨大潜力。

关键词：
AI模型；区块链；同态加密；数据安全
doi：10.12045/j.issn.1007-3043.2025.09.007
文章编号：1007-3043（2025）09-0038-05
中图分类号：TN915.08
文献标识码：A
开放科学（资源服务）标识码（OSID）：

Abstract：
The training and optimization of AI models highly rely on massive data. The sources of these data exhibit significant characteristics of diversity and dispersion. Traditional data sharing mechanisms are confronted with problems such as data leakage, data tampering, and privacy protection. Under this background, building an AI model data security sharing system becomes an urgent matter. This system relies on blockchain and homomorphic encryption technology to ensure that training data does not need to leave the local area. Through alliance collaboration training, it can achieve AI model development. In this way, data security sharing can be realized on a larger scale, thereby fully releasing the huge potential of AI models.

Keywords：
AI model；Blockchain；Homomorphic encryption；Data security

引用格式：程筱彪，张曼君，谢泽铖. 基于区块链的AI模型数据安全共享方案研究[J]. 邮电设计技术，2025（9）：38-42.

0 引言

人工智能（AI）技术的迅猛发展高度依赖大量高质量数据，数据共享作为整合多方数据资源的有效手段，能够大幅提升AI模型的准确性与泛化能力。然而，AI模型数据常涉及敏感信息[1]，如个人隐私数据、企业商业机密等。传统数据共享方式在数据安全与隐私保护方面存在诸多漏洞[2]，例如2023年某医院病历数据库遭黑客攻击，导致10万条隐私数据泄露，直接经济损失超2 000万元，这些问题严重阻碍了数据的有效共享。

现有研究已尝试探索数据安全共享模式，早期方案多依赖集中式隐私计算[3]，但存在计算开销大、单点故障风险等问题。近年来，区块链与AI的融合成为研究热点：Google提出的联邦学习框架实现了隐私保护[4]，但缺乏对模型参数篡改的防御机制；全同态加密算法虽能支持任意计算[5]，但计算复杂度极高，难以适配AI模型的大规模训练；部分同态加密算法在浮点数计算上更适用于AI场景[6]，但现有方案未明确密钥分布式管理策略。

相比之下，本方案创新提出“区块链+横向联邦学习+同态加密”的三层融合架构，在支持跨机构协作训

收稿日期：2025-07-08

练的同时，通过智能合约实现数据权属的动态更新。

1 相关技术概述

1.1 区块链技术

区块链是一种分布式账本技术，通过去中心化信任机制、数据不可篡改、智能合约自动化等核心优势[7]，正在重塑金融、医疗等行业的运作模式[8]。本方案采用联盟链架构，其核心设计如下。

a）共识机制。采用实用拜占庭容错（PBFT）算法，节点数量控制在10个左右，区块生成周期为5 s，确保交易确认的高效性与容错能力。

b）智能合约框架。基于Hyperledger Fabric 2.4，采用Go语言编写Chaincode，支持链上链下数据分离存储。

c）安全机制。引入智能合约形式化验证工具（如ProVerif），通过预定义规则检测合约逻辑漏洞，并设置合约升级通道，支持漏洞修复后的平滑迭代。

1.2 AI模型数据特点

AI模型数据是人工智能模型训练、优化的基础，具有规模大、类型多样、数据复杂度高、序列关联性强、价值密度差异大以及敏感性强等特点[9]。许多AI应用需要融合多领域数据，通过对跨领域数据的分析和挖掘，发现不同领域之间的潜在关联和规律，而数据的安全性与隐私性则关乎数据所有者的权益及应用的合规性。

2 方案整体设计

2.1 总体架构

本方案架构主要分为数据层、区块链层、智能合约层和应用层（见图1）。

a）数据层。负责数据的基础处理与安全保障，包括确定数据权属、追溯数据来源、原始数据预处理以及加密敏感数据。采用同态加密算法对原始数据进行加密，密钥通过分布式密钥生成协议生成，由3个以上可信节点共同持有。

b）区块链层。提供分布式账本技术，实现数据存储、共享与聚合，支持联盟协作训练并记录模型参数，保障数据的不可篡改和各参与方的协同训练。基于联盟链架构，节点包括数据提供方、监管节点以及聚合节点（负责模型参数聚合）。区块结构新增“操作类型”字段，明确区分“数据上链”“模型参数更新”“权限变更”等行为，支持细粒度追溯。

c）智能合约层。通过智能合约对数据访问权限、共享策略进行配置与管理，同时验证隐私保护情况。该层包含权限控制合约、聚合规则合约、审计合约3类合约。权限控制合约基于属性基访问控制（ABAC）模型，定义访问主体属性、资源属性及访问条件（如“仅用于模型训练，不可导出原始数据”）；聚合规则合约预设联邦平均（FedAvg）算法的参数（如学习率、迭代轮次上限）。

d）应用层。直接面向AI模型数据的提供者和使用者，为用户和外部系统提供数据交互接口、权限管理、共享记录展示、共享申请以及审计监管等功能，是用户与系统交互的入口。

2.2 核心功能

2.2.1 数据确权与溯源

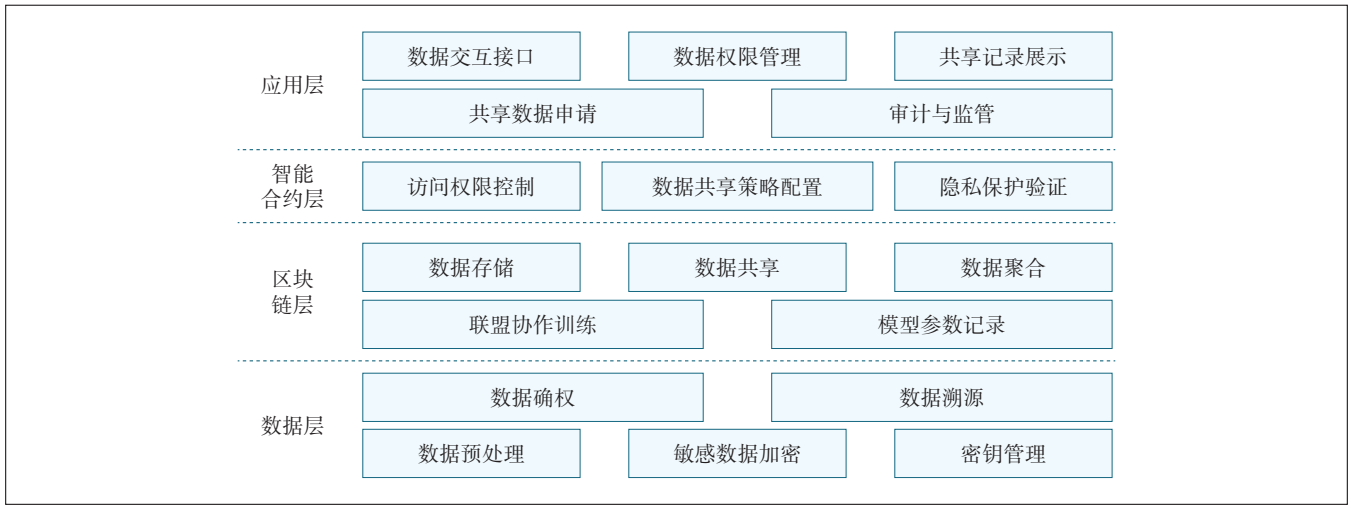

图1 AI模型数据安全共享系统架构

数据提供者将数据加密后,基于加密数据的摘要值、提供者的身份标识、时间戳生成该数据的唯一标识,作为该数据在系统中的编码,从而明确该数据的归属权。数据唯一标识的生成规则为:

数据标识=SHA-256(加密数据哈希+提供者DID+时间戳+数据类型标签)

其中,提供者DID基于W3C DID标准生成,与区块链地址绑定,确保身份的不可伪造;时间戳采用区块链节点的共识时间(避免本地时间篡改);数据类型标签(如"影像数据""文本数据")用于后续需求匹配。

将标识上链存储,并通过共识算法进行同步,可靠记录数据的确权信息,实现数据权属在全网的共识。同时,权属规则会同步到相应的智能合约中,当涉及数据交易、使用等操作时,智能合约自动验证参与者是否具备相应权限,保障数据所有者的权益。

在数据的整个生命周期中,相关信息都将被记录到区块链上,每个数据操作记录都被打上时间戳,形成完整的时间链条,精确反映数据的流转顺序和时间节点。当需要对数据进行溯源时,可通过应用层的查询接口,依据数据的标识、时间等信息,获取其完整的流转历史。

2.2.2 联盟协作训练

各方在本地基于加密数据进行模型训练,得到本地模型参数,然后利用联邦学习技术,在不泄露原始数据的情况下,对本地模型参数进行聚合。聚合后的全局模型参数被广播回各个参与节点,各方更新本地模型,继续下一轮训练。

每一轮训练结束后,利用区块链的共识机制对训练结果进行验证,确保结果的正确性和一致性。同时,将训练过程中的关键信息,如每轮的模型参数、训练时间、参与方贡献等记录在区块链上,方便后续的审计和追溯。

2.2.2.1 同态加密算法选择

本方案选用加法同态加密算法(Paillier算法)对数据进行加密处理。Paillier算法是一种加法同态加密算法,具有加法同态的特性,能够满足AI模型数据共享与协作训练的需求。

a)加密过程。数据提供者首先对原始数据m进行加密,选择随机数r,计算密文:

$$c = g^m r^n \bmod n^2$$

其中g和n是算法的公钥参数。加密后的数据c将被上传至区块链进行存储与共享。

b)同态运算。在联邦学习过程中,各参与节点对本地模型参数进行加密后上传至区块链。聚合节点利用同态加法特性对加密参数进行聚合计算。例如,对于2个加密参数c_1和c_2,聚合节点计算$c_1 \cdot c_2 \bmod n^2$,得到的结果对应原始参数之和的加密值。

2.2.2.2 横向联邦学习模式

横向联邦学习模式适用于同类型数据跨机构协作,如多医院的病历数据训练疾病预测模型,具体流程如下。

a)初始模型由联盟链上的可信机构发布,模型结构(如CNN、LSTM)及初始参数哈希上链存证。

b)本地训练时,参与方采用SGD优化器,批次大小(batch size)设为32,每完成10个批次计算一次本地梯度,并通过同态加密后上传。

c)聚合节点采用加权平均算法聚合梯度(权重与参与方数据量正相关),聚合结果的哈希值上链后,各参与方下载并验证(对比本地计算的聚合结果哈希与链上值,不一致则拒绝更新)。

d)训练终止条件。当连续3轮全局模型的准确率提升小于0.1%,或迭代次数达到50轮时,智能合约触发训练终止流程。

2.2.3 模型参数记录

在协作训练模型过程中,当满足特定条件时,如完成一轮训练、达到一定的准确率阈值等,智能合约将模型的参数,如权重、偏置、超参数等,以及每次训练迭代的结果、梯度信息等,作为交易记录存储在区块链的区块中。

参与者在每轮训练结束后也会将本地模型参数(或梯度)计算哈希值(如SHA-256),将哈希值与训练轮次、参与者身份(DID)、时间戳等元数据绑定,写入区块链。

模型参数上链采用"增量存储"策略,即仅记录每轮训练的参数变化量(而非完整参数),降低链上存储开销。参数元数据包括:

参数ID=SHA-256(模型ID+轮次号+聚合节点签名)

其中,模型ID与初始模型上链时的标识绑定,轮次号由智能合约自动递增,确保参数序列的连续性。

2.2.4 数据共享策略配置

系统管理员或者数据提供者在区块链上通过智能合约编程语言编写数据访问权限规则,实现细粒度的数据访问权限管理,防止未授权访问。规则可包括允许访问的主体(如特定组织、个人的区块链地址)、

访问条件（身份认证要求）、访问期限（访问时间、访问次数）等。规则支持动态更新，数据提供者可通过“策略更新交易”发起规则修改，当获得超2/3联盟节点的签名确认时，修改方可生效。

数据需求用户在应用层进行共享数据申请，详细说明所需数据的用途、使用期限、预期应用场景等信息。智能合约接收到请求后，依据预设的权限规则自动进行审批，如验证请求方身份是否符合要求等。若请求符合所有规则，智能合约批准请求，并对相关数据的访问合约进行更新。

2.3 数据安全共享整体流程

2.3.1 数据预处理与上链存证

数据提供用户对原始数据进行预处理，采用同态加密算法对数据加密，然后使用摘要算法计算加密数据的哈希值，最后生成该段数据的唯一标识，将加密后的数据、描述数据详情的元数据以及标识上链存储进行数据确权，同时将数据权属、共享策略规则同步到相应智能合约中。

2.3.2 需求匹配与智能合约触发

数据需求用户通过共享数据申请模块提交模型训练所需数据的各项指标，包括所需数据类型、模型目标、资源约束。智能合约对需求用户的身份认证信息进行验证，验证通过后，按照访问权限以及共享策略规则匹配对应的数据，生成协作训练任务，包括训练参与用户标识、参与训练数据标识、训练的各项参数。

2.3.3 联邦学习与安全训练具体流程

参与训练用户根据合约下载初始模型，在本地基于加密数据进行模型训练，得到本地模型参数，训练结果参数采用同态加密传输给参与节点，聚合节点（由智能合约指定）收到各参与用户的结果后利用同态加密算法直接对加密梯度进行聚合运算，计算其哈希值并上链。各参与方下载聚合后的参数，对比链上哈希值，验证完整性。重复本地训练—加密上传—链上聚合的流程，直至模型收敛或达到预设轮次。

若某参与方上传的参数与历史训练趋势偏差超过3σ（通过智能合约的异常检测模块判定），则该参数被标记为“可疑”，不参与本轮聚合，同时记录该节点的异常行为。

2.3.4 数据使用记录与追溯流程

区块链实时记录数据的每一次访问与训练操作，包括访问时间、访问者区块链地址、数据使用方式（如训练模型的算法、参数等）、训练过程中的模型参数、数据处理结果的哈希值等信息。当出现数据安全问题或需要审计时，监管机构或数据所有者可通过区块链上的记录，精确追溯数据的使用全过程，明确责任主体。

2.4 方案核心优势

2.4.1 数据共享与可信性

区块链可以为AI模型提供可信的数据共享机制，确保数据来源的真实性和可追溯性[10]。通过智能合约，可自动化执行数据共享过程，同时保护数据不被滥用。

2.4.2 模型训练与推理的可验证性

区块链记录AI模型的训练数据、参数调整和输出结果，确保训练过程的透明性和可验证性[11]，防止模型被恶意篡改。

2.4.3 去中心化与分布式计算

区块链支持去中心化的AI模型开发和共享[12]，允许不同机构在不共享原始数据的情况下协作训练模型。

3 方案场景验证

为验证本方案在实际场景中的可行性，选取多中心疾病预测模型训练作为实验场景，3家医院作为数据提供方与训练参与方，对比传统集中训练方案和本方案的各项训练数据。实验结果与分析如下。

a）数据隐私保护验证。网络抓包与本地存储检查显示，所有节点的原始数据均未以明文形式传输，模型参数在传输过程中始终以密文形式存在，仅授权聚合节点可解密聚合。

b）模型性能比对。本方案训练的模型准确率与集中式训练的仅相差0.3%（传统集中式的准确率为89.2%，本方案的准确率为88.9%），收敛轮次接近（传统集中式平均为32轮，本方案平均为34轮）。这表明本方案在保护数据隐私的同时，模型性能未受显著影响；且增加参与节点数量后，模型准确率更接近集中式训练结果，体现了多中心协作的优势。

c）防篡改与可追溯性结果。在防篡改测试中，恶意节点篡改参数记录后，区块哈希值发生变化，与前区块哈希无法匹配，系统触发异常告警，验证了区块链的防篡改能力；追溯测试显示，所有模型参数记录均可完整追溯至对应的训练轮次、参与节点DID及数据标识，溯源成功率100%，满足医疗监管审计要求。

d)效率分析。在区块链性能方面,联盟链的平均TPS为28.3,平均延迟为1.2 s,满足医疗场景的低频高安全需求;在训练耗时分布方面,本地计算占比65%,加密/解密占比20%,区块链共识占比15%。

4 方案安全与性能分析

4.1 安全性分析

a)数据安全方面。用户的数据均存储到本地,只有训练后的结果参数经过加密后共享到聚合节点[13],有效防止数据在传输以及存储阶段被攻击。

b)权限控制方面。智能合约在用户接入后会对认证信息进行验证,并对用户能够使用申请的数据以及使用条件进行细粒度的控制,降低了数据滥用风险。

c)操作溯源方面。数据的所属、数据的使用、模型的训练过程均会通过区块链进行记录,所有操作都有不可篡改的记录,数据一旦上链便难以被篡改,任何对数据的操作都会被完整记录并可追溯,确保数据的真实性和完整性。

4.2 性能分析

a)效率方面。链上存储用户标识、数据标识、操作日志等数据,这些数据的量级远小于训练数据本身,因此对区块链的数据写入速度要求不高,不会造成数据上链延迟。数据查询展示相关的操作对处理速率的要求不高,不会影响区块链的正常运行。智能合约实现自动化的权限管理和审批流程,减少人工干预,提高数据共享的效率。

b)可靠性方面。区块链的分布式架构使数据在多个节点上存储和备份,不存在单点故障问题[14]。即使部分节点出现故障或离线,也不会出现数据丢失或无法访问的问题。

c)资源开销方面。虽然区块链记录与加密操作会带来一定的存储与计算开销,但通过合理的数据存储策略以及分布式计算资源的利用,可在可接受范围内平衡性能与安全需求。

d)合规性方面。数据确权以及操作的全过程上链有助于满足法律法规对数据共享的合规要求,保障数据提供用户和数据需求用户的权益[15]。

5 结语

本方案根据AI模型数据的特点提出共享架构与流程,实现数据在安全、可控条件下的高效共享。随着区块链技术的不断发展与完善,以及AI应用场景的日益丰富,本系统具有广阔的发展空间。未来可进一步研究同态加密在高维AI数据(如图像、视频)中的计算效率瓶颈及优化思路,联盟链节点动态加入、退出时的共识机制适应性(如弹性共识算法设计)以及跨链场景下的数据共享兼容性(如与其他区块链系统的互操作方案)。

参考文献:

[1] 姜鑫,王春东.基于区块链的医疗数据安全共享模型研究[J].天津理工大学学报,2025,41(1):51-56.

[2] 滕亮,陈兵,赵开斌,等.基于区块链的医疗数据安全共享模型研究与应用[J].信息安全研究,2023,9(9):884-891.

[3] 李懿,王劲松,张洪玮.基于区块链与函数加密的隐私数据安全共享模型研究[J].大数据,2022,8(5):33-44.

[4] 夏虎,田雯,高建彬,等.基于区块链的外包安全多方统计计算可验证隐私保护方案[J].无线电工程,2024,54(4):835-847.

[5] 熊维,王海洋,唐祎飞,等.安全多方计算应用的隐私度量方法[J].信息安全研究,2024,10(1):6-11.

[6] 朱诗生,李朝清,黄仁俊,等.基于区块链的医疗数据安全共享模型与机制[J].计算机技术与发展,2020,30(10):123-130.

[7] 臧洪睿,杨婷婷,刘洪波,等.面向物联网的分布式联邦学习加密验证研究[J].计算机科学,2024,51(增刊1):1050-1054.

[8] 程雪婷,王玮茹,暴悦爽,等.基于联邦学习的多源异构数据安全融合方法[J].通信技术,2023,56(10):1173-1183.

[9] 邵航,高思琪,钟离,等.同态加密在隐私计算中的应用综述[J].信息通信技术与政策,2022(8):75-88.

[10] 刘嘉微,马兆丰,王姝爽,等.基于区块链的隐私信用数据受限共享技术研究[J].信息网络安全,2022,22(5):54-63.

[11] 何雨一,孙新芳,李晓冲,等.AI数据安全交易系统设计与实现[J].计算机时代,2025(6):49-55.

[12] 李辉.基于区块链技术的不动产多源异构数据安全共享模型研究[J].科技创新与应用,2025,15(10):95-98.

[13] 范瑞龙,章恒,马鹏,等.大模型数据共享中的安全风险评估与防范技术[J].数字技术与应用,2025,43(4):195-197.

[14] 魏群,王泽林,龙青良,等.面向下一代网络演进的新型网络共享技术[J/OL].电信科学:1-17(2025-06-19)[2025-06-23].https://link.cnki.net/urlid/11.2103.tn.20250618.1655.004.

[15] 王鹏,张弘,张玉,等.大数据与人工智能时代下数据安全的风险及应对措施[J].科学技术创新,2025(15):75-78.

作者简介:

程筱彪,副高级工程师,硕士,主要从事网络与信息安全研究工作;张曼君,正高级工程师,博士,主要从事网络与信息安全研究工作;谢泽铖,工程师,硕士,主要从事网络与信息安全研究工作。