

Design and Implementation of a Blockchain-Based Trustworthy AI Big Data Sharing Platform

Lu Peng

Shijiazhuang Linhe Technology Co., Ltd. Shijiazhuang City Hebei Province, 050000;

Abstract: With the rapid development of artificial intelligence and big data technologies, the demand for data sharing is increasing day by day. However, data security and trust issues have become the key factors hindering the efficient sharing of data. This paper proposes a design and implementation scheme for an AI big data trusted sharing platform based on blockchain. This platform utilizes the decentralized, immutable, and consensus mechanism characteristics of blockchain to ensure the security and credibility of data during the sharing process. Through smart contracts, data authorization access and transactions are realized, and combined with AI technology, data analysis and mining are conducted to enhance the utilization value of data. The overall architecture, functional module design, and key technology implementation of the platform are elaborated in detail. Through experimental verification, this platform can effectively solve the trust issues in data sharing and promote the efficient and trusted sharing of AI big data, which has important theoretical and practical significance.

Keywords: Blockchain; AI Big Data; Trusted Sharing Platform; Smart Contract; Data Security

DOI: 10.64216/3106-4620.26.01.011

In the current digital era, artificial intelligence (AI) and big data technologies have been widely applied in various fields, providing a significant impetus for social development and economic growth. AI technologies rely on a large amount of data for training and optimization to enhance their intelligence and decision-making capabilities; big data provides rich materials and insights for AI. However, the dispersion and closed nature of data have led to the emergence of data silos, making it difficult to achieve efficient data sharing and restricting the further development of AI and big data technologies.

The traditional data sharing methods have many problems, such as unclear data ownership, high risk of data tampering, and difficulties in data privacy protection, which seriously affect the credibility and security of data sharing. Blockchain technology, as an emerging distributed ledger technology, has decentralized, immutable, and consensus mechanism characteristics, providing new ideas and methods to solve the trust issues in data sharing.

1 Overall Architecture Design of the Blockchain-based AI Big Data Trust Sharing Platform

1.1 Architecture Overview

The overall architecture of the blockchain-based AI big data trust sharing platform mainly includes the data layer, blockchain layer, middleware layer, and application layer. The data layer is responsible for collecting, storing, and managing various types of big data, including structured and unstructured data. The blockchain layer is the core of the platform, using blockchain's distributed ledger technology to ensure the security and immutability of data, and achieving trust and collaboration among nodes through the consensus mechanism.

The middleware layer provides a series of interfaces and services for connecting the data layer and the blockchain layer, implementing functions such as data encryption, decryption, and authorized access. The application layer provides various application scenarios for users, such as data sharing, data analysis, and data transactions, and realizes data management and transactions through smart contracts.

This layered architecture design enables the platform to have good scalability and flexibility, able to adapt to different types and scales of data sharing needs.

1.2 Design of the Data Layer

The data layer is the data foundation of the platform, mainly including data sources, data collection module, and data storage module. Data sources can be various types of databases, file systems, sensors, etc., and the data collection module collects data from data sources and stores it in the platform. The data collection module can adopt batch collection and real-time collection methods, choosing the appropriate collection method according to different data sources and application requirements.

The data storage module is responsible for storing and managing the collected data, using a combination of distributed file systems and databases to ensure the reliability and scalability of data. To ensure data security, data is encrypted, and only authorized users can access and use the data.

The data layer also provides data cleaning and preprocessing functions, cleaning, transforming, and integrating the collected data to improve data quality and usability.

1.3 Design of the Blockchain Layer

The blockchain layer is the core part of the platform, adopting a consortium chain architecture, consisting of multiple nodes. Each node has a complete copy of the blockchain ledger, and through the consensus mechanism, it achieves trust and collaboration among nodes. The consensus mechanism uses the Practical Byzantine Fault Tolerance (PBFT) algorithm to ensure that the system can still operate normally in the event of partial node failures or malicious attacks.

The blockchain layer also provides smart contract functionality, using smart contracts to achieve authorized access and transactions of

data. Smart contracts are automatically executed contracts that execute corresponding operations when the preset conditions are met. In the data sharing platform, smart contracts can be used to define data access permissions, transaction rules, and fee settlement, ensuring the legal use of data and the fairness of transactions. To enhance the performance and scalability of the blockchain, the sidechain technology is adopted to separate some business processes from the main chain, thereby reducing the burden on the main chain. The data of the blockchain is compressed and optimized to reduce the occupation of storage space and transmission bandwidth.

2 Function Module Design of the AI Big Data Trusted Sharing Platform Based on Blockchain

2.1 Data Management Module

The data management module is responsible for the comprehensive management of data in the platform, including data registration, review, classification, and annotation. Data providers can register their data on the platform and submit an application for review. Reviewers review the submitted data to ensure its legality, accuracy, and completeness.

The reviewed data is classified and annotated to facilitate users' query and retrieval. The data management module also provides a data version management function, recording the update history and changes of the data to ensure its traceability.

To ensure data security, the data management module encrypts the data and sets different access permissions for different data. Only authorized users can access and use the corresponding data.

2.2 Data Sharing Module

The data sharing module is one of the core functions of the platform, responsible for implementing data sharing and exchange. Data providers can set their data as shared status and specify the scope and conditions of sharing. Data users can query and apply for shared data according to their own needs, and after being authorized by the data provider, they can obtain the corresponding data.

The data sharing module uses smart contracts to achieve automated data sharing and transactions. Smart contracts define the rules and conditions for data sharing, and when these conditions are met, the data sharing and transaction operations are automatically executed. To ensure the security and privacy of data, encryption technology is used to encrypt the data during the data sharing process.

The data sharing module also provides data sharing statistics and analysis functions, which statistically and analyze the data sharing situation and provide decision support for data providers and users.

2.3 Data Analysis Module

The data analysis module uses AI technology to analyze and mine the shared data, extracting valuable information and knowledge. This module provides various data analysis algorithms and tools, such as machine learning, deep learning, and data mining, and users can choose the appropriate algorithm and tool for data analysis according to their own needs.

The data analysis module also provides visualization functions, presenting the analysis results in the form of charts and reports to facilitate users' understanding and decision-making. To improve the efficiency and accuracy of data analysis, data preprocessing and feature extraction are carried out to reduce data noise and redundancy.

The data analysis module also supports distributed computing and parallel processing, using cluster computing resources to improve the speed and performance of data analysis.

3 Key Technology Implementation of the AI Big Data Trusted Sharing Platform Based on Blockchain

3.1 Blockchain Technology Implementation

In the implementation of blockchain technology, Hyperledger Fabric is adopted as the underlying framework of the consortium chain. Hyperledger Fabric is an open-source enterprise-level blockchain platform with high scalability, flexibility, and security. It adopts a modular design, allowing users to select different consensus mechanisms, encryption algorithms, and smart contract languages according to their own needs.

In the blockchain network, nodes are divided into sorting nodes and peer nodes. Sorting nodes are responsible for sorting and packaging transactions to form blocks; peer nodes are responsible for storing and verifying blocks and executing smart contracts. The PBFT consensus mechanism ensures consistency and reliability among nodes. To enhance the performance of the blockchain, the chaincode technology is adopted to optimize the smart contracts. Chaincode is a program that runs on the blockchain network and can implement complex business logic and data processing. The data of the blockchain is compressed and indexed to reduce storage space and query time.

3.2 Implementation of Smart Contract Technology

Smart contracts are written in the Solidity language, which is a high-level programming language specifically used for Ethereum smart contracts. In the platform, smart contracts are mainly used to implement functions such as authorized access to data, transaction rules, and fee settlement.

The deployment and execution of smart contracts require strict testing and verification to ensure their correctness and security. When deploying smart contracts, they are uploaded to the blockchain network and verified and agreed upon by nodes. When executing smart contracts, corresponding operations are automatically triggered based on preset conditions.

To ensure the security of smart contracts, formal verification technology is used to verify smart contracts, checking for vulnerabilities and security risks. The execution process of smart contracts is monitored and audited to ensure legality and compliance.

3.3 Implementation of Data Encryption and Privacy Protection Technology

In terms of data encryption, a combination of symmetric and asymmetric encryption methods is adopted. Symmetric encryption algorithms such as AES are used to encrypt large amounts of data, while asymmetric encryption algorithms such as RSA are used to encrypt and sign keys. Data is encrypted during storage and transmission to ensure the security and privacy of the data.

To protect users' privacy, homomorphic encryption and zero-knowledge proof technologies are used. Homomorphic encryption allows calculations to be performed on encrypted data without decrypting it, thereby protecting the privacy of the data. Zero-knowledge proof technology allows the prover to prove the authenticity of a statement without revealing any additional information, used for verifying the identity and permissions of users.

Establish a data access control mechanism to strictly manage and audit the access behavior of users. Only authorized users can access and use the corresponding data, and their access behavior is recorded and monitored.

4 Experiment and Evaluation of the Blockchain-based AI Big Data Trusted Sharing Platform

4.1 Setup of Experimental Environment

To verify the performance and functions of the blockchain-based AI big data trusted sharing platform, an experimental environment was set up. The experimental environment includes a blockchain network composed of multiple nodes, as well as data collection, storage, and analysis systems. Nodes are simulated using virtual machines and run on different servers.

The data collection system uses the Kafka message queue to collect various types of data, while the data storage system uses the Hadoop distributed file system and MySQL database for storage. The data analysis system uses the Python programming language and the TensorFlow deep learning framework for data analysis and mining.

In the experimental environment, different parameters and scenarios were set up to simulate different scales and complexity of data sharing and analysis tasks.

4.2 Performance Evaluation Indicators and Methods

To evaluate the performance of the platform, several indicators were adopted: throughput, response time, data security, and data availability. Throughput refers to the number of transactions processed by the platform per unit time, response time refers to the time it takes for users to initiate a request and receive a response, data security refers to whether the data is tampered with and leaked during sharing, and data availability refers to whether the data can be accessed when needed.

The platform was evaluated using a combination of stress testing and functional testing. Stress testing simulates a large number of concurrent requests to test the performance of the platform under high load conditions; Functional testing conducts individual tests on all the functions of the platform to ensure its normal operation.

4.3 Experimental Results and Analysis

The experimental results show that the AI big data trusted sharing platform based on blockchain has a high throughput and low response time. Even under high load conditions, the platform can maintain stable performance and handle a large number of concurrent requests. The data security and availability of the platform have been effectively guaranteed, and no tampering or leakage of data occurred during the sharing process.

By conducting experiments on different scales and complexity of data sharing and analysis tasks, it is found that the platform has good scalability and adaptability. With the increase in data volume and the increase in task complexity, the performance of the platform can still meet the actual needs.

Compared with traditional data sharing platforms, the AI big data trusted sharing platform based on blockchain has obvious advantages in data security, credibility, and traceability. It can effectively solve the trust issues in data sharing and promote the efficient and trustworthy sharing of AI big data.

5 Conclusion

This paper designs and implements a blockchain-based AI big data trusted sharing platform. Through the decentralized, immutable and consensus mechanism characteristics of blockchain, the security and credibility of data during the sharing process are ensured. The platform adopts a layered architecture design, including the data layer, blockchain layer, middleware layer and application layer, which has good scalability and flexibility.

The functional module design of the platform is elaborated in detail, including the data management module, data sharing module and data analysis module, as well as the key technology implementation, such as blockchain technology, smart contract technology and data encryption and privacy protection technology. Through experiments, the platform has high throughput, low response time and good data security and availability, which can effectively solve the trust issues in data sharing and promote the efficient and trusted sharing of AI big data.

However, this platform still has some shortcomings, such as the performance bottleneck of blockchain and the security issues of smart contracts. The future research direction will focus on further optimizing the performance of blockchain, improving the security and reliability of smart contracts, and expanding the application scenarios and functions of the platform. Strengthening the integration with other fields, such as the Internet of Things and cloud computing, to promote the development of AI big data trusted sharing.

References

-
- [1] Dai Weiqi, Li Ming, Zhao Kexuan, et al. Blockchain Marketing Label Trading System for E-commerce Alliances [J]. Computer Research and Development, 2025, 62(1): 269-280. DOI: 10.7544/issn1000-1239.202330217.
- [2] Zhou Yuanjian, Qin Dongmei, Liu Yining, et al. Design of Trustworthy Warehouse Document System Based on Blockchain [J]. Information Security and Network Protection, 2019, (6): 84-90. DOI: 10.3969/j.issn.1671-1122.2019.06.011.
- [3] Liu Zhanghui, Lin Zheyu, Chen Hanlin, et al. Data Trustworthy Sharing Scheme Based on Consortium Blockchain [J]. Computer Science, 2025, 52(11): 398-407. DOI: 10.11896/jsjcx.241000169.
- [4] Cao Xian, Zhang Heng, Gao Xu, et al. Research on the Spatial Big Data Platform of Smart City Based on Blockchain [J]. Planner, 2020, 36(24): 46-51.
- [5] Xi Yanana, Zhang Hongyu, Gao Xin, et al. Big Data Knowledge Sharing Model of Energy Internet Based on Blockchain [J]. Electric Power Construction, 2022, 43(3): 123-130. DOI: 10.12204/j.issn.1000-7229.2022.03.013.

Author Profile: Lu Peng (1987.2.9 -) Male, Han ethnicity, native place: Nan Gong, Hebei Province, Education: Bachelor's degree, Professional title: Senior Engineer, Research Direction: Mechanical and Electrical Engineering and Computer AI Big Data