DECENTRALIZED & COLLABORATIVE AI ON BLOCKCHAIN

Praveena Muvva, Jahnavi Kotapati, Thanuja Nalajala

Instructor: Dr. Pooria Yaghini

Abstract

The proposed framework aims to decentralize machine learning models by leveraging blockchain technology for collaborative dataset building and model hosting. Unlike centralized systems, which often limit access to proprietary datasets and charge for model predictions, this framework allows for freely accessible and continuously updated models shared on a public blockchain. To incentivize participation and maintain model accuracy, the framework integrates both financial and gamified incentive structures. Notably, existing proposals like DInEMMo and DanKu offer limited access or lack continuous updating capabilities, highlighting the novelty of this approach in addressing the current centralization of artificial intelligence. The framework's modularity enables the implementation of various incentive mechanisms to encourage participation while discouraging manipulation.

Background and Proposed Framework

The framework facilitates collaborative dataset creation and model hosting through smart contracts, with an Incentive Mechanism validating transactions. The Data Handler stores data and metadata on the blockchain, while the Machine Learning Model allows continuous updates and prediction queries. Users are incentivized with monetary payments or virtual "karma" points, promoting decentralized dataset building and participation.

Integrating machine learning models into decentralized networks, focusing on Ethereum's blockchain. It introduces smart contracts and proposes staking deposits for compliance and simpler incentives. Sections detail ML data handling, incentives, implementation, blockchain use, issue resolution, and future, applicable beyond Ethereum.

Incentive Mechanisms Overview

Gamification: Incentivizes data contributions through points and badges, like Stack Exchange, based on metrics like quantity, diversity, and submission frequency.

Rewards Mechanism Based on Prediction Markets: Offers monetary rewards for accurate data contributions, with phases including commitment, participation, and reward, while preventing manipulation through cryptographic schemes.

Deposit, Refund, and Take: Self-Assessment: Validates data submissions using a deposit-refund system, where contributors reclaim funds for valid data and penalties for incorrect submissions.

Benefits of Blockchain

Using a blockchain like Ethereum for hosting and deploying machine learning models offers several advantages over traditional methods. Firstly, it provides a public, persistent, and decentralized platform, ensuring model availability and trust through visibility across decentralized nodes. Secondly, versioning becomes simpler, allowing easy reverting to previous model versions by updating to specific blockchain blocks, aiding in version control and performance evaluation. Thirdly, smart contracts enable model evolution over time, adapting to changing environmental conditions such as novel concepts. Lastly, blockchain-based smart contracts offer transparency and trust, unlike closed systems, ensuring confidence in model updates and contributor compensation.

Potential Issues

The potential issues with the proposed blockchain-based model training framework include the risk of submitting bad or ambiguous data, potential network congestion, and the requirement for deposits, which may deter new contributors. To address these, the framework suggests implementing robust incentive mechanisms, emphasizing unambiguous data submission, and considering third-party interfaces to manage deposits and rewards.

Future Work and Conclusion

Future work includes researching the effectiveness of unsupervised models like clustering and generative models for outlier detection and data generation tasks. Overcoming challenges with complex models, such as cost constraints and data corruption, will involve pre-computation strategies and recovery methods. Tailoring incentive mechanisms for different model types is crucial for maximizing contributor engagement. Privacy concerns can be addressed by submitting model updates instead of raw data directly to the blockchain. Overall, the framework aims to enhance scalability and tackle emerging challenges in machine learning deployment.