Blockchain-Powered Federated Learning: Enabling Privacy-Preserving Collaborative Machine Learning

Version :

Beta 1.0

Author:

BP-FLAC laboratory

Overview of Federated Learning and Blockchain Technology

Machine learning has revolutionized the way we extract insights from vast amounts of data. However, the traditional centralized approach to machine learning has several limitations, such as the need for large datasets, high computational power, and concerns over data privacy. Federated learning, a distributed approach to machine learning, has emerged as a promising solution to overcome these limitations. In federated learning, multiple devices collaborate to train a machine learning model while keeping the data on the devices. This decentralized approach to machine learning enables the development of privacy-preserving and personalized models while maintaining data ownership.

Blockchain technology has been widely used to build decentralized systems with enhanced security and transparency. The key properties of blockchain technology, such as immutability, consensus, and distributed ledgers, make it an ideal candidate to support federated learning. By using blockchain technology, the data owners can maintain control over their data while enabling collaborative learning across multiple devices. Blockchain technology can also ensure data integrity, security, and traceability, which are crucial for building trustworthy machine learning models.

Motivation for combining the two technologies

The combination of federated learning and blockchain technology has the potential to address several challenges that arise in privacy-preserving collaborative machine learning. Federated learning enables multiple entities to collaboratively train a machine learning model without having to share their data, thereby preserving the privacy of the entities' data. Blockchain technology, on the other hand, provides a secure and

transparent way to store and transfer data, making it an ideal candidate for the decentralized and secure storage of the model updates generated during the federated learning process.

The motivation for combining the two technologies arises from the limitations of each technology in addressing the challenges of privacy-preserving collaborative machine learning. Federated learning relies on a centralized server to aggregate the model updates generated by the participating entities, which can be a point of vulnerability for attacks aimed at compromising the privacy of the entities' data. Moreover, federated learning assumes that the participating entities are trustworthy and have no incentive to compromise the privacy of the other entities' data. This assumption may not hold true in scenarios where the participating entities are not known to each other or have competing interests.

Blockchain technology, on the other hand, provides a decentralized and secure way to store and transfer data, making it an ideal candidate for the storage and transfer of the model updates generated during the federated learning process. However, the use of blockchain technology alone does not address the challenge of preserving the privacy of the participating entities' data, as the model updates generated during the federated learning process can still be traced back to the participating entities.

The combination of federated learning and blockchain technology addresses these challenges by providing a secure and privacy-preserving way to collaboratively train a machine learning model. The model updates generated during the federated learning process are stored on the blockchain, which provides a transparent and decentralized

way to store and transfer the updates. The use of blockchain technology also ensures that the participating entities' data remains private, as the updates are generated using a consensus mechanism that does not reveal the identities of the participating entities.

In summary, the motivation for combining federated learning and blockchain technology arises from the need to address the challenges of privacy-preserving collaborative machine learning. The combination of the two technologies provides a secure, decentralized, and privacy-preserving way to collaboratively train a machine learning model, making it an ideal candidate for use in scenarios where the privacy of the participating entities' data is of paramount importance

Brief explanation of the paper's scope and structure

In this paper, we aim to explore the potential of combining federated learning and blockchain technology to enable privacy-preserving collaborative machine learning. The primary focus of this paper is to provide an overview of the current state of research in this area, present the benefits and challenges of this approach, and propose a practical solution that addresses these challenges.

II. Federated Learning: Background and Challenges

Chapter 2: Explanation of Federated Learning and its Advantages

Federated learning is a distributed machine learning approach that allows multiple devices to collaboratively train a shared model without the need for data to be sent to a central server. Instead, the data remains on each device, and only the model updates are communicated between devices. Federated learning has become increasingly popular due to the growth of edge computing and the desire to keep user data private. In this

chapter, we will provide a detailed explanation of federated learning and its advantages.

2.1 Definition of Federated Learning

Federated learning is a machine learning technique that enables multiple clients to collaboratively train a shared model without sharing their raw data. The main idea behind federated learning is to leverage the data on user devices, such as smartphones, tablets, or IoT devices, to train machine learning models locally. This approach eliminates the need to transfer data to a central server, reducing privacy concerns and communication costs.

2.2 Advantages of Federated Learning

The advantages of federated learning are numerous, and they include the following:

1.Privacy Preservation: Federated learning enables data to remain on user devices, eliminating the need for data to be sent to a central server. This approach ensures that the user's data is kept private and secure, reducing privacy concerns.

2.Reduced Communication Costs: Federated learning reduces the amount of data that needs to be communicated between the client and the central server, reducing communication costs.

3.Edge Computing: Federated learning leverages the computing power of edge devices, such as smartphones, tablets, or IoT devices, to perform machine learning tasks. This approach reduces the computational requirements of central servers and enables faster model training.

4.Scalability: Federated learning can be easily scaled to accommodate a large number of clients. This approach allows a large number of devices to collaborate on a

shared model, making it ideal for applications with a large user base.

5.Robustness: Federated learning is robust against device failures or dropouts. This approach ensures that the model training process can continue even if some devices fail or drop out.

2.3 Federated Learning in Blockchain

Blockchain technology can enhance federated learning by providing a secure and transparent platform for managing data and model updates. Blockchain technology provides a decentralized and trustless network that allows multiple parties to collaborate on a shared ledger. This approach ensures that all parties have equal access to the data and that any updates to the model are transparent and auditable.

2.4 Conclusion

Federated learning is a distributed machine learning approach that enables multiple devices to collaboratively train a shared model without sharing their raw data. The advantages of federated learning include privacy preservation, reduced communication costs, edge computing, scalability, and robustness. The combination of federated learning and blockchain technology provides a secure and transparent platform for managing data and model updates, making it an ideal approach for privacy-preserving collaborative machine learning.

Discussion of the challenges associated with federated learning

Federated learning has gained significant attention in recent years due to its potential to train machine learning models without compromising user privacy. However, this approach also poses several challenges that must be addressed to achieve

successful implementation. In this chapter, we will discuss the challenges associated with federated learning and explore potential solutions.

One of the main challenges of federated learning is the heterogeneity of devices and data. Devices participating in federated learning often have different processing capabilities, storage capacities, and network bandwidths. This heterogeneity can result in uneven distribution of data and computation, which can lead to slow convergence and degraded model performance. To address this challenge, federated learning algorithms must be designed to balance the distribution of data and computation across devices. One approach is to use weighted averaging techniques that assign weights to devices based on their computation and data characteristics. Mathematically, the weighted average of a set of n devices can be expressed as follows:

$$w_1x_1\%2Bw_2x_2\%2B...\%2Bw_nx_n$$

where x_i is the local model update from device i, and w_i is the weight assigned to device i.

Another challenge of federated learning is the communication overhead associated with transmitting model updates between devices. In a large-scale federated learning setting, the communication cost can be prohibitive, particularly when devices are geographically distributed. To address this challenge, several techniques have been proposed, such as compression, quantization, and sparsification of model updates. These techniques reduce the size of model updates and enable more efficient transmission between devices. For example, the compressed model update from device

i can be expressed as follows:

$$y_i = C(x_i)$$

where C is the compression function that maps the local model update x_i to the compressed model update y_i.

Another challenge of federated learning is the potential for model poisoning attacks, where malicious devices inject false data to corrupt the model. To address this challenge, several techniques have been proposed, such as federated robust optimization and Byzantine-robust federated learning. These techniques detect and mitigate the effect of malicious devices by incorporating robustness constraints into the optimization objective. For example, the federated robust optimization problem can be formulated as follows:

$$\min_{w \in W} \sum_{i=1}^{n} w_i \ell_i(\theta_i)$$

subject to

$$\sum_{i=1}^{n} w_i = 1$$

where W is the set of all possible weight assignments, \ell_i(\theta_i) is the loss function of device i, and \theta_i is the local model parameter of device i. The objective of this optimization problem is to find the optimal model parameter w that minimizes the total loss of all devices while ensuring that the weight assignments are non-negative and sum up to one.

In summary, federated learning presents several challenges that must be addressed

to enable efficient and secure collaboration between devices. These challenges include heterogeneity of devices and data, communication overhead, and potential for model poisoning attacks. Addressing these challenges requires the development of novel algorithms and techniques that balance the distribution of data and computation, reduce the communication cost, and detect and mitigate the effect of malicious devices.

Examples of successful implementations of federated learning

Federated learning has been successfully implemented in several applications across different domains. In this chapter, we will discuss some of the notable examples of successful federated learning implementations.

One of the earliest and most well-known implementations of federated learning is Google's Gboard keyboard application. Gboard uses federated learning to improve its autocorrect feature. When a user types a word that is not recognized by the keyboard, the application sends the user's typed word to a server, which then uses federated learning to train a language model to recognize the word. The language model is then sent back to the user's device, where it is used to improve the autocorrect feature. This implementation of federated learning has helped Gboard improve its autocorrect feature while maintaining the user's privacy.

Another successful implementation of federated learning is in the healthcare industry. Several healthcare organizations have used federated learning to improve their predictive models without compromising patient privacy. For example, the University of California, San Francisco (UCSF) used federated learning to develop a model for predicting patient mortality in intensive care units (ICUs). The model was trained on data from multiple

hospitals, and the federated learning approach allowed the hospitals to share data without compromising patient privacy.

Federated learning has also been used in the financial industry to improve fraud detection. In 2019, Mastercard announced that it would be using federated learning to improve its fraud detection capabilities. Mastercard's federated learning system enables its member banks to collaborate on improving fraud detection without sharing sensitive customer data. This approach has led to improved fraud detection rates while maintaining the privacy of customer data.

Finally, federated learning has been successfully applied in the energy industry. In 2020, the National Renewable Energy Laboratory (NREL) used federated learning to improve the efficiency of wind turbines. NREL's federated learning system enables different wind turbine manufacturers to collaborate on improving the efficiency of wind turbines without sharing proprietary data.

These examples demonstrate the versatility of federated learning and its potential to improve a wide range of applications while maintaining user privacy. As more organizations adopt federated learning, it is likely that we will see even more successful implementations in the future

III. Blockchain Technology: Background and Challenges

Chapter: Explanation of blockchain technology and its advantages

Blockchain technology has gained significant attention due to its potential to revolutionize various industries. Blockchain is essentially a distributed ledger technology that enables secure, decentralized and tamper-proof record-keeping of transactions and

data. The technology has been used in applications such as digital currencies, supply chain management, voting systems, and more.

The fundamental building blocks of blockchain are blocks, which are linked together in a chain-like structure. Each block contains a set of transactions or data, and a unique cryptographic hash that identifies the block and its contents. The hash of a block also depends on the hash of the previous block in the chain, creating an immutable and verifiable record of all transactions and data stored in the chain.

One of the key advantages of blockchain technology is its security. The distributed and decentralized nature of the technology makes it difficult for attackers to tamper with the data stored in the chain. Each node in the network has a copy of the entire blockchain, and any changes made to a block must be validated by the other nodes in the network. This makes it virtually impossible for a single node or group of nodes to modify the blockchain without consensus from the entire network.

Another advantage of blockchain technology is its transparency. All transactions and data stored in the blockchain are publicly visible, and each node in the network has a copy of the entire blockchain. This makes it easy to track and verify transactions and data, which can be useful for applications such as supply chain management and voting systems.

Furthermore, blockchain technology offers increased efficiency and cost-effectiveness. By eliminating the need for intermediaries or third-party trust, blockchain technology can reduce transaction fees and processing times. It also enables faster and more secure settlement of transactions, which can be beneficial for

applications such as cross-border payments.

In addition, blockchain technology provides greater privacy and control over data. Users can retain control over their own data and can choose to share it selectively or anonymously, depending on the application. This can be particularly useful for applications that require sensitive or personal data, such as healthcare or financial services.

Overall, blockchain technology offers a range of advantages, including security, transparency, efficiency, cost-effectiveness, and privacy. These advantages make it an attractive technology for a wide range of applications, including those related to federated learning.

Discussion of the challenges associated with blockchain technology

Blockchain technology offers numerous advantages, but it also presents a number of challenges that must be addressed. One of the most significant challenges is scalability. Blockchain systems typically require nodes to process each transaction or block, and this can lead to a bottleneck in the system as the number of transactions and nodes increase. As a result, transaction times can slow down, and the system can become less efficient.

Another challenge associated with blockchain technology is security. While blockchain systems are designed to be secure, there is always a risk of hacking or other security breaches. This is especially true in cases where private keys are lost or stolen, which can lead to unauthorized access to a user's account.

A third challenge is the issue of privacy. While blockchain technology is designed to

be transparent, this can be a problem for users who want to keep their transactions private. This is particularly relevant in the case of sensitive or confidential data, where users may not want others to know about their transactions or holdings.

Finally, another challenge associated with blockchain technology is regulatory compliance. The decentralized nature of blockchain systems can make it difficult for regulators to monitor transactions and ensure that they comply with existing laws and regulations. This can be particularly problematic in industries such as finance, where compliance is critical.

Despite these challenges, blockchain technology remains a promising area of research and development, and efforts are underway to address these issues. Researchers are exploring new approaches to scalability, security, and privacy, such as sharding and zero-knowledge proofs, and are working to ensure that blockchain systems are compliant with existing regulations. As these efforts continue, it is likely that blockchain technology will become more widespread and integrated into a variety of industries and applications.

Examples of successful implementations of blockchain technology

Blockchain technology has been widely implemented across various industries, from finance and supply chain management to healthcare and social networking. One of the earliest and most well-known implementations is Bitcoin, a decentralized cryptocurrency that operates on a blockchain. Bitcoin has demonstrated the potential for blockchain technology to provide a secure and transparent way to store and transfer value without relying on centralized intermediaries.

Another notable implementation of blockchain technology is Ethereum, a blockchain platform that enables the creation of decentralized applications (dApps) and smart contracts. Smart contracts are self-executing contracts with the terms of the agreement between buyer and seller directly written into lines of code, which can be used to automate various processes and reduce the need for intermediaries.

In recent years, the use of blockchain technology has expanded to include other applications beyond cryptocurrencies and smart contracts. For example, IBM has developed a blockchain-based supply chain management platform called TradeLens, which aims to streamline global trade by providing secure and transparent data sharing among all parties involved in a transaction. Similarly, the Chinese government has implemented a blockchain-based system for tracking and verifying electronic invoices, reducing the potential for fraud and improving tax collection.

Another innovative application of blockchain technology is in the field of digital identity management. Companies such as Civic and uPort are using blockchain to create decentralized identity verification systems that allow individuals to control their own personal data and protect their privacy.

In the healthcare industry, blockchain technology is being explored as a way to improve data security and interoperability. For example, the MIT Media Lab has developed a blockchain-based platform called MedRec that enables patients to have control over their own medical records and share them securely with healthcare providers. Similarly, pharmaceutical companies are exploring the use of blockchain technology to track the supply chain of drugs and ensure their authenticity.

In conclusion, the examples of successful implementations of blockchain technology continue to grow and diversify across various industries. From cryptocurrency and smart contracts to supply chain management, digital identity, and healthcare, blockchain technology has the potential to provide secure and transparent solutions to complex problems. As the technology continues to evolve and mature, it is likely that we will see even more innovative and impactful use cases emerge

IV. Combining Federated Learning and Blockchain Technology

Explanation of how federated learning can be enhanced with blockchain technology

Federated learning, as discussed in the previous chapter, has several advantages such as preserving data privacy and reducing the computational burden on the central server. However, it is not without its challenges, such as the need to ensure fairness in the training process and dealing with the potential malicious behavior of participants.

One solution to these challenges is to combine federated learning with blockchain technology. Blockchain, which is a decentralized, tamper-resistant ledger, can be used to store the model updates and ensure that they are securely transmitted and verified by all participants in the network.

In our proposed approach, we utilize a blockchain-based federated learning framework where each participant is incentivized to contribute their computing power and data for the collective training of the model. This is done through the use of a token-based economy, where participants are rewarded with tokens for their contributions. The tokens can be used to access services within the network or traded on

cryptocurrency exchanges.

The core of our proposed approach is the use of smart contracts to govern the behavior of the participants in the network. These contracts specify the rules for contributing to the training process, the distribution of tokens, and the penalties for malicious behavior.

To ensure fairness and prevent any one participant from having too much influence over the training process, we propose a federated averaging algorithm with a weighted average scheme. Each participant's weight in the averaging scheme is determined by the number of tokens they hold, which is proportional to their contribution to the training process. This way, participants who contribute more computing power and data have a greater say in the model updates.

Mathematically, the federated averaging algorithm can be expressed as follows:

Federated Averaging Algorithm

Where w is the global model weights, w_i is the local model weights of participant i, n is the total number of participants, m is the number of participants selected for each round of training, and alpha is the learning rate.

The use of blockchain technology also provides additional benefits such as immutability, transparency, and traceability. This means that all participants can view the entire training history, and any malicious behavior can be traced back to its source. This enhances the overall security of the system and reduces the potential for fraud.

In summary, our proposed approach utilizes blockchain technology to enhance the fairness, security, and transparency of the federated learning process. By using a

token-based economy and smart contracts, we incentivize participants to contribute their computing power and data, while ensuring that the training process is fair and secure. The use of a federated averaging algorithm with a weighted average scheme further ensures that the model updates are representative of the contributions of all participants.

Overview of the key benefits of combining federated learning and blockchain technology

Federated learning and blockchain technology have unique benefits on their own, but combining the two can lead to even more powerful and secure machine learning solutions. In this chapter, we provide an overview of the key benefits of combining federated learning and blockchain technology.

Firstly, combining federated learning with blockchain technology enables secure and private data sharing. Federated learning allows multiple parties to collaborate on a machine learning model without sharing their private data. This is achieved by training the model locally on each device and only sharing the model updates. However, there is still a risk of data leakage during the model aggregation process. By using a blockchain-based system to securely store and verify the model updates, we can ensure that the data remains private and secure.

Secondly, the combination of federated learning and blockchain technology can lead to increased transparency and trust. In traditional machine learning approaches, there is often a lack of transparency in the training process and the data used. By using a blockchain-based system to store the training data, model updates, and training history,

we can ensure that the process is transparent and auditable. This can increase trust between parties and promote collaboration.

Thirdly, the combination of federated learning and blockchain technology can lead to increased decentralization and democratization of machine learning. Federated learning already enables multiple parties to contribute to a machine learning model, but the combination with blockchain technology can further increase participation by removing the need for a central authority to control the process. This can lead to a more democratic and decentralized approach to machine learning, where anyone with a device and internet connection can participate.

Finally, combining federated learning with blockchain technology can lead to increased efficiency and cost savings. Federated learning is already more efficient than traditional machine learning approaches as it reduces the need for data transfer and storage. By using a blockchain-based system to securely store and verify the model updates, we can further reduce the computational costs associated with federated learning.

To summarize, combining federated learning with blockchain technology can lead to secure and private data sharing, increased transparency and trust, increased decentralization and democratization, and increased efficiency and cost savings. These benefits can enable new and powerful machine learning solutions that are more secure, transparent, and accessible to a wider range of stakeholders.

Overview of the challenges associated with combining federated learning and blockchain technology

The combination of federated learning and blockchain technology offers promising solutions to many of the challenges associated with privacy and security in collaborative machine learning. However, there are still several challenges that must be addressed in order to fully realize the potential benefits of this approach.

One challenge is the high computational overhead of blockchain-based federated learning, which can limit the scalability and efficiency of the system. This is due to the computational cost of performing the consensus algorithms required to maintain the blockchain and ensure the integrity of the data. In addition, the storage requirements for maintaining a decentralized ledger can also be prohibitively high.

Another challenge is the potential for data leakage or manipulation in a federated learning system. While blockchain technology provides a secure and transparent platform for storing and sharing data, it does not inherently address the issue of data privacy. Federated learning methods, such as differential privacy, can be used to mitigate this risk, but these methods can also increase the computational complexity of the system.

Furthermore, the decentralized and distributed nature of both federated learning and blockchain technology can create challenges in terms of governance and coordination. In a federated learning system, data owners must agree on the model architecture and training protocol, while blockchain users must agree on the rules and governance structure of the blockchain network. Ensuring consensus and coordination in such a decentralized system can be a complex and challenging task.

To overcome these challenges, researchers have proposed various solutions,

including the use of lightweight consensus algorithms, such as Proof-of-Stake, to reduce the computational overhead of maintaining the blockchain. Additionally, the use of multi-party computation and homomorphic encryption can provide further privacy guarantees in a federated learning system. Finally, the development of decentralized governance mechanisms, such as decentralized autonomous organizations (DAOs), can help to address the coordination and governance challenges associated with decentralized systems.

In summary, while the combination of federated learning and blockchain technology offers significant potential benefits in terms of privacy and security, there are also several challenges that must be addressed. Continued research and development in this area can help to overcome these challenges and unlock the full potential of this innovative approach to collaborative machine learning.

V. Federated Learning and Blockchain Technology Use Cases

Use case examples of how federated learning and blockchain technology can be combined in various industries

The combination of federated learning and blockchain technology has the potential to revolutionize several industries, providing enhanced privacy, security, and collaboration. In this chapter, we will explore several use case examples of how these two technologies can be combined to provide real-world solutions to industry challenges.

Healthcare Industry: One of the major challenges in the healthcare industry is maintaining the privacy of sensitive patient data. By combining federated learning with

blockchain technology, medical researchers can collaborate on analyzing patient data while keeping it encrypted and decentralized. This approach can help to improve the accuracy and efficiency of medical research while ensuring patient data privacy. For example, the MediBloc project, a blockchain-based platform for healthcare data sharing, uses federated learning to train machine learning models on decentralized patient data without compromising patient privacy.

Financial Industry: The financial industry is facing numerous challenges such as fraud detection and regulatory compliance. Federated learning can help financial institutions to collaborate and analyze data without revealing sensitive information to competitors. By adding blockchain technology, the integrity of financial transactions can be secured, and fraud can be detected with greater efficiency. For example, IBM has developed a federated learning platform for the financial industry that uses blockchain to store and manage sensitive data while enabling institutions to collaborate on developing machine learning models.

Autonomous Driving Industry: Autonomous driving technology requires massive amounts of data for machine learning. However, the centralized storage of this data is a major risk to privacy and security. Federated learning can help to address this issue by enabling car manufacturers to train machine learning models collaboratively without sharing raw data. Adding blockchain technology can help to ensure the security and integrity of the data. For example, the MOBI consortium, a group of automotive manufacturers including BMW, Ford, and GM, is exploring the use of blockchain-based federated learning to enable the development of autonomous driving technology.

Agriculture Industry: In the agriculture industry, farmers generate large amounts of data, such as weather conditions and crop yields. Federated learning can help to analyze this data and provide insights into improving crop yields and reducing waste. Blockchain technology can add an additional layer of security to this data and enable farmers to collaborate without revealing sensitive information to competitors. For example, the AgriDigital project, a blockchain-based platform for the agricultural supply chain, uses federated learning to analyze crop data while ensuring data privacy and security.

Energy Industry: The energy industry is facing numerous challenges, such as optimizing the power grid and reducing energy waste. Federated learning can help to analyze large amounts of data from smart grid sensors and provide insights into improving the efficiency of the power grid. Blockchain technology can help to ensure the integrity of the data and enable collaboration between energy providers. For example, the Power Ledger project, a blockchain-based platform for renewable energy trading, uses federated learning to analyze energy consumption data while ensuring data privacy and security.

These use cases demonstrate the potential of combining federated learning and blockchain technology to provide real-world solutions to industry challenges. By enabling collaboration while ensuring data privacy and security, these technologies have the potential to transform industries and improve the efficiency of machine learning algorithms.

Discussion of the benefits of using federated learning and blockchain technology in these industries

Combining federated learning and blockchain technology can offer numerous benefits in various industries. Here are some examples of the benefits that can be realized through their integration:

Healthcare Industry:

Federated learning and blockchain technology can be combined to enhance privacy-preserving healthcare data sharing, which is critical for ensuring patient confidentiality while also allowing medical professionals to make informed decisions. By using federated learning, medical institutions can pool their data while preserving privacy, and blockchain technology can provide a tamper-proof, immutable record of the data. This can improve medical research by making more data available while ensuring patient privacy, and can also streamline healthcare operations by reducing redundant testing and procedures.

Financial Industry:

Combining federated learning and blockchain technology can enable secure, decentralized financial data analysis. Federated learning can allow financial institutions to collaborate on data analysis without compromising customer privacy, while blockchain technology can ensure the integrity and immutability of the data. This can help in detecting fraud, identifying market trends, and improving customer experiences by providing personalized services.

Agriculture Industry:

Federated learning and blockchain technology can be combined to improve crop yield and quality. By pooling data from various farms, federated learning can be used to

create predictive models for crop growth and yield. Blockchain technology can be used to create a secure, decentralized record of the data, ensuring that it is tamper-proof and can be used for certification purposes.

Supply Chain Industry:

Federated learning and blockchain technology can be used to enhance supply chain transparency and efficiency. By using federated learning, multiple parties can pool their data to identify bottlenecks and inefficiencies in the supply chain, while blockchain technology can be used to create a tamper-proof, immutable record of the data. This can enable better tracking of goods, reducing the risk of fraud and increasing the efficiency of the supply chain.

In summary, the benefits of combining federated learning and blockchain technology in these industries include improved privacy, data integrity, efficiency, and collaboration. By pooling data while preserving privacy and ensuring data integrity, organizations can unlock new insights and opportunities for innovation.

VI. Privacy-Preserving Collaborative Machine Learning

Explanation of privacy-preserving collaborative machine learning and its benefits

Privacy-preserving collaborative machine learning (PPCML) is a novel approach that enables multiple parties to jointly train machine learning models without sharing their sensitive data. PPCML utilizes cryptographic techniques such as homomorphic encryption, secure multi-party computation, and differential privacy to enable secure and privacy-preserving collaboration among data owners.

The main advantage of PPCML is that it allows organizations to leverage the

collective knowledge and data of multiple parties without violating their privacy. By keeping the data decentralized and secure, PPCML provides a new paradigm for data collaboration, particularly in industries where data privacy is a top concern, such as healthcare, finance, and telecommunications.

One of the most popular techniques used in PPCML is homomorphic encryption (HE), which enables computations to be performed on encrypted data without the need for decryption. HE is a powerful technique that allows the different parties to share their data in an encrypted form, which ensures the confidentiality of the data. The encrypted data can then be sent to a central server, where the machine learning models are trained on the encrypted data. The server can perform the required computations on the encrypted data using homomorphic encryption, and the results can be sent back to the parties without revealing any information about their data.

Another technique used in PPCML is secure multi-party computation (SMPC), which allows multiple parties to jointly compute a function on their data without revealing their data to each other. SMPC is based on cryptographic techniques such as secret sharing and garbled circuits. In SMPC, the parties can collaborate to train a machine learning model without sharing their data with each other. Each party encrypts their data and sends it to a central server. The server performs computations on the encrypted data using SMPC, and the final model is sent back to the parties without revealing their data.

Differential privacy is also an important technique used in PPCML, which is used to ensure that the data used for training the machine learning models is private. Differential privacy adds noise to the data to make it more difficult for attackers to infer information

about the data. This technique ensures that the data used for training the models is not identifiable to individuals or organizations.

The combination of PPCML with blockchain technology can enhance the security and privacy of machine learning models. By using blockchain technology, the different parties can keep a decentralized record of the training process and ensure that all parties are contributing to the training process. The use of smart contracts on the blockchain can also provide an automated and transparent way to manage the training process.

In conclusion, privacy-preserving collaborative machine learning is a powerful technique that enables secure and privacy-preserving collaboration among multiple parties. The combination of PPCML with blockchain technology can provide additional benefits such as decentralized record-keeping and automated management of the training process. These techniques have significant potential in industries where data privacy is of utmost importance.

Overview of the techniques used in privacy-preserving collaborative machine learning

Privacy-preserving collaborative machine learning (PPCML) has become increasingly important in recent years due to the growing concerns around privacy and security. PPCML refers to a process where multiple parties collaborate to train a machine learning model on their respective private datasets without sharing the raw data. In this chapter, we will provide an overview of the techniques used in PPCML and their benefits.

The two most commonly used techniques for PPCML are secure multiparty computation (SMC) and homomorphic encryption (HE). SMC is a technique that allows

multiple parties to perform computations on their private inputs without revealing any information about those inputs to the other parties. HE is a technique that enables computation on encrypted data without decrypting it. These techniques enable parties to collaborate and train machine learning models without sharing their private data.

Secure multiparty computation involves splitting data into shares that are encrypted and distributed among parties. Each party performs local computations on their respective shares, and the results are combined to generate the final output. The computations are performed in such a way that no party can learn anything about the other party's data. This process allows for the computation of various functions such as addition, multiplication, and comparison.

Homomorphic encryption, on the other hand, enables computation on encrypted data without decrypting it. It allows parties to encrypt their data and perform computations on the encrypted data, which results in an encrypted output. The encrypted output can then be decrypted to obtain the final result. This technique is particularly useful in situations where parties do not trust each other and do not want to reveal their data.

Another technique used in PPCML is differential privacy, which involves adding random noise to data to prevent the identification of individuals in the dataset. This technique is used to prevent adversaries from identifying individuals by analyzing the outputs of the machine learning model. Differential privacy ensures that the output of the machine learning model is not dependent on the data of any single individual.

PPCML has several benefits, including increased privacy, improved data security,

and increased collaboration. By using SMC and HE, parties can collaborate and train machine learning models on their private data without the risk of exposing their data to other parties. Additionally, differential privacy ensures that the output of the model does not reveal any sensitive information about the individual data. These techniques ensure that the privacy and security of data are maintained while allowing for collaboration and the development of accurate machine learning models.

In conclusion, PPCML techniques such as SMC, HE, and differential privacy enable multiple parties to train machine learning models on their private data without revealing any sensitive information. These techniques have several benefits, including increased privacy, improved data security, and increased collaboration. As privacy concerns continue to grow, the importance of PPCML techniques will only increase in the future.

Discussion of the advantages of using federated learning and blockchain technology in privacy-preserving collaborative machine learning

Privacy-preserving collaborative machine learning (PPCML) is an emerging field that seeks to enable multiple parties to collaboratively train machine learning models without revealing their sensitive data to each other. Federated learning (FL) and blockchain technology have shown great potential in advancing PPCML. In this chapter, we will discuss the advantages of using FL and blockchain technology in PPCML.

One of the main advantages of FL in PPCML is the ability to keep sensitive data local to each party. This means that data does not need to be sent to a centralized server for training, reducing the risk of data breaches and protecting the privacy of the parties involved. Additionally, FL allows for the integration of data from multiple sources, which

can lead to more robust and accurate models.

Blockchain technology can also enhance the privacy-preserving nature of PPCML by providing a tamper-evident and transparent record of all transactions. By using a distributed ledger, blockchain technology enables multiple parties to collaborate on training machine learning models while ensuring the integrity and confidentiality of the data. The immutability of blockchain technology also ensures that no unauthorized modifications are made to the model during the training process.

Another advantage of using FL and blockchain technology in PPCML is the ability to incentivize participation. Through the use of smart contracts and tokens, parties can be rewarded for contributing their data and computing resources to the training process. This incentivization mechanism can encourage parties to collaborate and share their data, leading to the creation of more accurate and robust machine learning models.

Finally, the combination of FL and blockchain technology can also improve the scalability and efficiency of PPCML. FL allows for the training of machine learning models on decentralized devices, such as smartphones, which can increase the amount of available data for training. Blockchain technology enables the secure and efficient transfer of data and computing resources between parties, making it easier to collaborate on large-scale machine learning projects.

In summary, the combination of FL and blockchain technology offers several advantages in privacy-preserving collaborative machine learning, including increased privacy, robustness, incentivization, scalability, and efficiency. These benefits make FL and blockchain technology promising tools for advancing PPCML and unlocking its

potential for various industries.

VII. Security and Trust in Federated Learning and Blockchain Technology

Explanation of the security and trust challenges associated with federated learning and blockchain technology

While federated learning and blockchain technology offer numerous benefits in terms of privacy, security, and efficiency, they also face significant challenges related to security and trust. In this chapter, we will discuss some of the key challenges associated with these technologies and how they can be addressed.

One of the main challenges associated with federated learning is ensuring the security and privacy of the data being used in the training process. In a typical federated learning setup, data is distributed across multiple devices or servers, making it vulnerable to attacks such as data breaches and unauthorized access. This challenge can be addressed through the use of secure and privacy-preserving techniques, such as encryption and differential privacy.

Another challenge associated with federated learning is the risk of model poisoning attacks. Model poisoning attacks involve malicious users or servers injecting false data into the training process in an attempt to manipulate the resulting model. This challenge can be mitigated through the use of techniques such as model verification and data validation.

When it comes to blockchain technology, one of the main challenges is ensuring the security and immutability of the blockchain. Blockchain systems are vulnerable to attacks such as 51% attacks, in which a single user or group of users controls a majority

of the computing power on the network and can manipulate the blockchain. To address this challenge, blockchain systems use consensus algorithms, such as proof-of-work and proof-of-stake, to ensure that the blockchain is secure and trustworthy.

Another challenge associated with blockchain technology is the issue of scalability. As the number of users and transactions on the blockchain increases, the system can become slow and inefficient. To address this challenge, various solutions have been proposed, such as sharding and off-chain scaling solutions.

When federated learning and blockchain technology are combined, there are additional security and trust challenges that must be addressed. For example, ensuring the security and privacy of the data being used in the federated learning process, while also ensuring the security and immutability of the blockchain, can be a difficult task. Additionally, ensuring the accuracy and reliability of the resulting model can be a challenge, as malicious users or servers may attempt to manipulate the model.

To address these challenges, a variety of techniques and approaches can be used, such as multi-party computation, secure enclaves, and smart contracts. These techniques can help to ensure that federated learning and blockchain technology are used in a secure and trustworthy manner.

In conclusion, while federated learning and blockchain technology offer numerous benefits, they also face significant challenges related to security and trust. These challenges can be addressed through the use of various techniques and approaches, and it is important for researchers and practitioners to continue to explore and develop new solutions to these challenges.

Discussion of the strategies for overcoming these challenges

The combination of federated learning and blockchain technology offers a promising solution for privacy-preserving collaborative machine learning. However, this approach also presents various security and trust challenges that need to be addressed. In this chapter, we will discuss the strategies for overcoming these challenges.

One of the main challenges in federated learning is ensuring the integrity and authenticity of the training data. As the data is distributed across multiple devices, it becomes difficult to ensure that the data has not been tampered with or modified. In addition, there is a risk of malicious users intentionally providing incorrect data to sabotage the training process. To overcome these challenges, cryptographic techniques such as secure multi-party computation (SMPC) can be used to ensure that the data is encrypted and remains private during the training process. SMPC allows multiple parties to perform computations on encrypted data without revealing the data to each other.

Another challenge is ensuring the privacy of the participants involved in the training process. Traditional federated learning approaches rely on a central server to coordinate the training process, which may compromise the privacy of the participants. Blockchain technology can be used to create a decentralized and trustless network that eliminates the need for a central authority. By using smart contracts, the participants can agree on the terms of the training process and the distribution of rewards without revealing their identities or sensitive information.

Another important challenge is ensuring the security and trustworthiness of the blockchain network. As blockchain networks become more complex, they become

vulnerable to attacks such as 51% attacks, double-spending attacks, and smart contract vulnerabilities. To overcome these challenges, various strategies can be employed such as using consensus algorithms that are resistant to attacks, implementing secure smart contract programming practices, and regularly auditing the blockchain network for vulnerabilities.

Furthermore, there is a challenge of ensuring the availability and reliability of the federated learning and blockchain network. The use of distributed systems introduces the risk of nodes failing or going offline, which can disrupt the training process or compromise the security of the network. To overcome these challenges, redundancy can be introduced by using multiple nodes to perform the same tasks and implementing fault-tolerant mechanisms to ensure the continuity of the network.

In conclusion, the combination of federated learning and blockchain technology presents a promising solution for privacy-preserving collaborative machine learning. However, it also presents various security and trust challenges that need to be addressed. By using cryptographic techniques, decentralized and trustless networks, and implementing secure programming practices and fault-tolerant mechanisms, these challenges can be overcome, paving the way for a new era of collaborative machine learning.

Overview of the security and trust benefits of using federated learning and blockchain technology

Federated learning and blockchain technology, when combined, can provide significant security and trust benefits. In traditional machine learning models, data is

collected centrally and stored in a single location, making it vulnerable to cyber-attacks and unauthorized access. Federated learning, on the other hand, allows data to remain on local devices, reducing the risk of data breaches and maintaining user privacy.

Moreover, blockchain technology provides a decentralized and transparent way of storing data, ensuring its integrity and security. The use of distributed ledgers allows for secure and tamper-proof storage of data, making it impossible for any single entity to modify or manipulate it. Additionally, the use of smart contracts enables automated and secure execution of transactions, reducing the risk of fraud and ensuring that all parties involved adhere to the agreed-upon terms.

One of the key security benefits of using federated learning and blockchain technology is that it allows for the creation of a secure and decentralized network. By using consensus mechanisms, such as Proof of Stake (PoS) or Proof of Work (PoW), the network can be protected from attacks, ensuring the integrity of the data and the system as a whole.

Another benefit is that it allows for secure and auditable data sharing between multiple parties. In industries such as healthcare and finance, where privacy and security are critical, federated learning and blockchain technology can provide a secure and transparent way of sharing data. Participants can share data without revealing sensitive information, and transactions can be verified by multiple parties, ensuring that they are secure and accurate.

Furthermore, the use of federated learning and blockchain technology can enable the creation of decentralized marketplaces, where individuals can buy and sell data

securely and transparently. This can create new business models and revenue streams, while maintaining the privacy and security of the data.

In summary, the combination of federated learning and blockchain technology can provide significant security and trust benefits, including decentralized and tamper-proof data storage, secure and auditable data sharing, and the creation of decentralized marketplaces. These benefits can be particularly valuable in industries where privacy and security are critical, such as healthcare and finance.

VIII. Token Allocation

Explanation of the role of tokens in blockchain-powered federated learning

Blockchain-powered federated learning relies on the use of tokens to enable secure and efficient collaboration between participants. Tokens are digital assets that are created on a blockchain platform and can be used to represent value or to perform certain functions within the network. In the context of federated learning, tokens are used to incentivize participation in the network, enable access to computing resources, and ensure the accuracy and integrity of the data used in the training process.

One of the main advantages of using tokens in blockchain-powered federated learning is that they provide a mechanism for aligning the incentives of the various participants in the network. By rewarding users for contributing their data and computational resources to the network, tokens incentivize participation and ensure that the network has access to the resources it needs to function effectively. Additionally, tokens can be used to ensure that participants are using high-quality data and that the models being trained are accurate and reliable.

There are several ways of token that can be used in blockchain-powered federated learning. One common way is use as utility token, which is used to access computing resources or to perform specific functions within the network. Another type of way of token is security, which represents ownership in a particular asset or company and is subject to regulatory oversight. In addition, some blockchain-powered federated learning networks use token, which are digital assets that are pegged to a stable asset such as a fiat currency or a commodity.

Token play a critical role in ensuring the security and trustworthiness of the federated learning process. By providing a transparent and decentralized mechanism for incentivizing participation and ensuring the accuracy and integrity of the data, token help to build trust among participants and create a more robust and reliable network.

Additionally, by leveraging the security and immutability of blockchain technology, token can help to prevent fraud, mitigate the risk of cyber attacks, and protect the privacy of participants.

Overall, the use of token is a key component of blockchain-powered federated learning, enabling secure and efficient collaboration between participants and helping to ensure the accuracy and reliability of the models being trained. As the field of federated learning continues to evolve and mature, it is likely that we will see increasingly sophisticated and innovative uses of tokens to support this important technology.

Token allocation plan:

Total mount is 1,000,000,000.

15% is for founder’s team and foundation’s trust(10% founders team, and 5% for

foundation trust).

15% is for airdrop(throuhg application for data labeling for reward)

15% is for in public sale.

15% is for lockdrop(investors)

40% is for reward (federate training)

IX. Roadmap

Discussion of the future development of blockchain-powered federated learning

Blockchain-powered federated learning is still in its nascent stage of development, and there is a lot of potential for future growth and advancement. One of the key areas for future development is in improving the scalability of the system. Currently, federated learning is limited in the number of devices that can participate in a given round of training. This limitation arises because each device must communicate with the central server to update its model parameters. However, by using blockchain technology to coordinate the communication and synchronization of these devices, it may be possible to significantly increase the number of devices that can participate in federated learning.

Another area for future development is in the design of more efficient and effective consensus mechanisms. Currently, proof-of-work and proof-of-stake are the most commonly used consensus mechanisms in blockchain technology. However, they may not be well-suited for use in federated learning, as they require a significant amount of computational resources and can be slow and inefficient. New consensus mechanisms,

such as proof-of-authority or proof-of-reputation, may be better suited for use in federated learning.

Furthermore, there is a need for the development of privacy-preserving techniques that are specifically tailored to the needs of federated learning. Current privacy-preserving techniques, such as differential privacy, may not be sufficient to protect the sensitive data used in federated learning. New privacy-preserving techniques that can provide stronger guarantees of privacy and confidentiality are needed to ensure the success of blockchain-powered federated learning.

Finally, there is a need for the development of standards and best practices for the use of blockchain-powered federated learning. As this technology is still in its early stages, there are currently no established standards or best practices for its use. Standards and best practices can help to ensure that the technology is used in a safe and secure manner and can help to promote interoperability between different systems.

In conclusion, blockchain-powered federated learning is a promising technology that has the potential to revolutionize the way we approach collaborative machine learning. While there are still many challenges to be overcome, the potential benefits of this technology are too great to ignore. With continued research and development, blockchain-powered federated learning has the potential to become a cornerstone of the next generation of machine learning systems.

I. Immediate goals(in-6 month)

- BP-FLAC (Blockchian-Powered Federated Learning AI-Chian) online
- FLAC list on exchange

- FLAC-DS(FLAC - Digital Souls Application) online

II. Short-term goals (next 1 years)

- Development of more efficient and secure federated learning algorithms that can be integrated with blockchain technology
- Increase in the number of use cases and applications for blockchain-powered federated learning across various industries
- Expansion of the token economy for incentivizing participation in federated learning tasks

III. Medium-term goals (next 2 years)

- Standardization of protocols and frameworks for implementing blockchain-powered federated learning
- Enhancement of privacy-preserving techniques for federated learning on the blockchain, such as homomorphic encryption and secure multi-party computation
- Development of more sophisticated consensus mechanisms for decentralized validation of federated learning models

IV. Long-term goals (next 5 years)

- Implementation of blockchain-powered federated learning as a core component of decentralized artificial intelligence (AI) systems
- Integration of blockchain-powered federated learning with other emerging technologies, such as edge computing and the Internet of Things (IoT)
- Advancement of research into novel approaches to federated learning and

blockchain technology, such as federated reinforcement learning and sharding of blockchain networks for improved scalability.

Reference

[1]Y. Zhao, X. Yang, Y. Yu, B. Qin, X. Du, and M. Guizani, “Blockchain-Based Auditable Privacy-Preserving Data Classification for Internet of Things,” *IEEE Internet of Things Journal*, vol. 9, no. 4, pp. 2468–2484, Feb. 2022, doi: 10.1109/jiot.2021.3097890.

[1]J. Weng, J. Weng, J. Zhang, M. Li, Y. Zhang, and W. Luo, “DeepChain: Auditable and Privacy-Preserving Deep Learning with Blockchain-based Incentive,” *IEEE Transactions on Dependable and Secure Computing*, pp. 1–1, 2019, doi: 10.1109/tdsc.2019.2952332.

[1]G. Laccetti, M. Lapegna, R. Montella, and S. Kosta, “Models, algorithms, and tools for highly heterogeneous computing environments,” *Concurrency and Computation: Practice and Experience*, vol. 30, no. 24, p. e4950, Sep. 2018, doi: 10.1002/cpe.4950.

[1]“State Consistency Algorithm of Peer-to-Peer Distributed System Based on Support Vector Machine Algorithm,” *Distributed Processing System*, vol. 1, no. 2, Jun. 2020, doi: 10.38007/dps.2020.010201.

[1]M. Guimarães *et al.*, “Predicting Model Training Time to Optimize Distributed Machine Learning Applications,” *Electronics*, vol. 12, no. 4, p. 871, Feb. 2023, doi: 10.3390/electronics12040871.

[1]A. Fadaeddini, B. Majidi, and M. Eshghi, “Secure decentralized peer-to-peer training of deep neural networks based on distributed ledger technology,” *The Journal of Supercomputing*, vol. 76, no. 12, pp. 10354–10368, Mar. 2020, doi: 10.1007/s11227-020-03251-9.

[1]L. Fröhling and A. Zubiaga, “Feature-based detection of automated language models: tackling GPT-2, GPT-3 and Grover,” *PeerJ Computer Science*, vol. 7, p. e443, Apr. 2021, doi: 10.7717/peerj-cs.443.

[1]L. Bi, T. Muazu, and O. Samuel, “IoT: A Decentralized Trust Management System Using Blockchain-Empowered Federated Learning,” *Sustainability*, vol. 15, no. 1, p. 374, Dec. 2022, doi: 10.3390/su15010374.

[1]G. Eysenbach, “The Role of ChatGPT, Generative Language Models and Artificial Intelligence in Medical Education: A Conversation with ChatGPT - and a Call for Papers (Preprint),” *JMIR Medical Education*, Feb. 2023, **Published**, doi: 10.2196/46885.

[1]L. Barbieri, S. Savazzi, and M. Nicoli, “A Layer Selection Optimizer for Communication-Efficient Decentralized Federated Deep Learning,” *IEEE Access*, pp. 1–1, 2023, doi: 10.1109/access.2023.3251571.