IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 August 2023, accepted 28 August 2023, date of publication 4 September 2023, date of current version 11 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3311713

RESEARCH ARTICLE

Decentralized Machine Learning Governance: Overview, Opportunities, and Challenges

DANA ALSAGHEER[1], LEI XU[2], AND WEIDONG SHI[1], (Senior Member, IEEE)
[1]Department of Computer Science, University of Houston, Houston, TX 77004, USA
[2]Department of Computer Science, Kent State University, Kent, OH 44240, USA

Corresponding author: Dana Alsagheer (dralsagh@CougarNet.UH.EDU)

This work was supported by the University of Houston.

ABSTRACT Researchers have started to recognize the necessity for a well-defined ML governance framework based on the principle of decentralization and comprehensively defining its scope of research and practice due to the growth of machine learning (ML) research and applications in the real world and the success of blockchain-based technology. In this paper, we study decentralized ML governance, which includes ML value chain management, decentralized identity for the ML community, decentralized ownership and rights management of ML assets, community-based decision-making for the ML process, decentralized ML finance, and risk management.

INDEX TERMS Blockchain, DAO, decentralization, governance, MLOps.

I. INTRODUCTION

THE last decade, machine learning (ML) has created widespread interest around the globe for its potential to transform human society. The advance of ML-based technologies like deep learning has enabled a wide range of applications (e.g., speech translation-transcription, computer image understanding, speech generation, image generation, ML-generated software, protein structure predictions [1], online recommendations, industrial robot automation, financial asset management, cyber security defense). To support ML growth and adoption, researchers and practitioners have proposed the concept of ML governance to manage the interactions between ML stakeholders and the ML systems [2]. ML governance plays a crucial role in the long-term success of ML as a significant source of technology innovations to make the future world a better place to live. However, the existing discussion of ML governance is narrowly defined.

Powered by the success of blockchain technology, the decentralized governance model has become popular in managing a community of stakeholders without reliance on a central entity for decision-making [3], [4]. The decentralized governance framework has been successfully applied and validated in various real-world applications, including decentralized autonomous organizations (DAOs), decentralized finance (DeFi) governance, and governance of blockchain protocols. These real-world examples demonstrate the effectiveness and potential of decentralized governance in managing and governing various aspects of decentralized systems, such as medical applications [5].

In this work, we expand the concept of ML governance under the lens of decentralization. We proposed a new framework of decentralized ML governance encompassing ML value chain management, decentralized identity for the ML community, decentralized ownership and rights management of ML assets, decentralized decision-making for the ML process, decentralized ML finance, and decentralized ML risk management. Most of the ML governance concepts described in this paper are new in the literature. The work drastically expands the scope of ML governance. Introducing decentralization to ML governance opens new research topics like community-owned and community-managed ML processes and DeFi for ML. It facilitates the integration of ML governance with blockchain-based innovations. The combination of ML governance and decentralization will catalyze the further growth of ML and create a new frontier for decentralized

The associate editor coordinating the review of this manuscript and approving it for publication was Yuan Gao.

governance. To summarize, this paper makes the following main contributions: **(i)** We describe a new framework of ML governance based on the principle of decentralization and define its scope of research and practice; **(ii)** We discuss details of the decentralized ML governance framework and provide a comprehensive view of its components; **(iii)** We present research opportunities, challenges, and open problems in each area of the decentralized ML governance model to spur further research and thinking; **(iv)** We compare with related concepts such as MLOps and the prior work on ML governance to highlight the new contributions.

II. RELATED WORKS

In this section, we provide an overview of existing efforts that relate to the theme of this paper.

A. MACHINE LEARNING GOVERNANCE

With the growing complexity of ML applications, particularly in real-world use cases, researchers have started recognizing the need for a well-defined governance mechanism. For instance, ML governance is believed to be a prerequisite to establishing trust between stakeholders in ML development to achieve responsible ML applications [2]. ML governance is the comprehensive process of managing access, implementing policies, enforcing regulations, and monitoring activities related to machine learning models. It aims to balance the advantages of utilizing ML while considering the associated security and privacy risks. ML governance encompasses measures to ensure that ML models are used responsibly and securely, protecting sensitive data and mitigating potential risks. This involves establishing guidelines, protocols, and frameworks to control access to ML models, enforcing policies to ensure compliance with security and privacy regulations, and tracking and monitoring activities to maintain transparency and accountability throughout the ML lifecycle.

The work focuses on creating a security, privacy, and accountability governance foundation for ML systems to avoid model bias, inaccuracy, and other issues. Centralized governance faces a massive problem due to the growing Ml research. One limitation is the lack of transparency because the central authority controls the system. The system is more vulnerable to security risks since the central authority has complete access to all the data, which means models with high risks are hard to operate. Scalability, as the central node, can only handle a limited number of clients at a time. ML can leverage blockchain technology or smart contracts to provide a decentralized and secure framework due to several advantages regarding trust, transparency, and decentralized governance within ML workflows. Provenance information enhances trustfully and transparently in the resulting models and allows for auditable and reproducible ML processes. Moreover, smart contracts provide a secure data-sharing environment among multiple parties involved in ML projects by defining clear rules and conditions within the contract [6], [7], [8]. Using blockchain technology with database-stored procedures can be more complicated than smart contracts. However, stored procedures and smart contracts have strengths and limitations in implementing complex ML models.

B. MLOps

MLOps, a combination of ML (Machine Learning) and DevOps (Development Operations), encompasses best practices that integrate ML development and operations within a unified framework. It involves various components such as ML data sources and datasets, repositories of ML models and metadata, automated ML pipelines and workflow processes, and software ML artifacts like containers for training and hosting services [9]. MLOps aims to automate the lifecycle of machine learning algorithms, starting from model training to deployment and retraining with new data. By streamlining the ML development process, MLOps achieves economies of scale by leveraging consolidated hardware and software resources, similar to other areas where cloud-based infrastructure is utilized. Major cloud service providers and ML hardware vendors have embraced this concept, leading to a projected industry value of over USD 6 billion by 2028 [10]. Currently, the definition of MLOps mainly revolves around infrastructure-level tools and processes. However, support for ML governance, as described in [2] and [11], is still in its early stages of development.

C. CENTRALIZED VS. DECENTRALIZED MACHINE LEARNING

Centralized machine learning is collecting and storing data in a central location, where the data is processed to train machine learning models. The trained model is returned to the clients or deployed in the relevant application. However, this classical approach raises many concerns, such as privacy due to storing the data in a central location, and it may be vulnerable to attacks. Furthermore, training in machine learning and deep learning models requires massive computing resources [12]. Taking advantage of the growing trend of decentralized computing infrastructures, efforts exist to develop decentralized ML systems where data collection, model training, and inference can be deployed over a decentralized computing environment. Researchers are leveraging technologies such as federated machine learning or blockchain to achieve such a goal, where models can be trained and updated without a centralized entity to protect sensitive data such as medical data [13], [14], [15], [16]. Federated learning (FL) is introduced as a distributed ML approach. FL enables collaborative machine learning (ML) model training without data sharing [17].

However, the traditional Federated Learning (FL) system has limitations, including a single point of failure with the central aggregator, malicious clients and false data, and the lack of incentives for participants. Integrating blockchain technology into FL will significantly improve security by leveraging blockchain's decentralization, anonymity, and traceability. The central aggregator can be replaced with a

peer-to-peer blockchain system. This decentralization eliminates the single point of failure and allows blockchain nodes to handle global model aggregation. This integration enhances the reliability and efficiency of the FL system while incentivizing participants to contribute honestly and provide reliable data [18].

It is worth pointing out that the concept of decentralized ML governance is different from distributed ML. Distributed ML may involve centralized or decentralized governance. These two concepts distinguish from one another. In the case of decentralized ML governance, the governance framework is proposed to apply to the governance, management, and control issues that may exist in all types of ML approaches. For distributed ML like federated ML, swarm learning, or gossip learning, due to their unique learning environments and assumptions, it may be easier to adopt certain aspects of decentralized governance aligned with the recent advances in research integrating blockchains with distributed ML techniques such as federated ML.

Decentralized machine learning (ML) presents a fertile ground for research but also brings various challenges from both research and practical perspectives. These challenges include concerns related to privacy and security. For example, there is a risk of sensitive data leakage to untrusted resource contributors during model training. Additionally, there is a potential for security risks arising from malicious participants in the system, such as attacks on training data and models and tampering with the ML system by adversaries due to its openness to public participants. Performance limitations, when compared to cloud and data center-based ML, cost concerns, and difficulties in audit and compliance guarantee are also prevalent in decentralized ML [19].

To address these performance limitations, LedgerDB offers purpose-built features that enhance throughput, audibility, and data management. These features are designed explicitly for permissioned blockchains. LedgerDB employs the TSA two-way peg protocol to improve security and ensure the ability to remove outdated records, enhancing storage efficiency and compliance with regulatory requirements [20], [21]. Notably, LedgerDB focuses exclusively on permissioned blockchains rather than public blockchains. Integrating blockchain into distributed learning models, particularly in medical imaging, increases transparency and accelerates AI adoption in multicentric studies. It eliminates the reliance on a single server for computations while maintaining performance levels comparable to centralized approaches [16]. It is essential to highlight that decentralized ML governance as a conceptual framework does not mandate how it should be realized. Depending on a targeted scenario, it may be achieved using DAOs, blockchains, distributed ledgers, or other blockchain-like systems. This article focuses on defining the scope of decentralized ML governance and analyzing research questions, challenges, and future research directions. For a specific application, it is up to the practitioners to decide the implantation details best suited for the needs, like types of concrete infrastructures, smart contract languages, etc.

D. CENTRALIZED VS. DECENTRALIZED GOVERNANCE

Centralized governance is a traditional hierarchical organization governance model in which a central authority or a few decision-makers control decision-making power. Centralized governance has various disadvantages, such as the conflict of interests between managers and shareholders and the lack of transparency [22]. Recognizing various limitations of centralized governance, in recent years, people have started experimenting with decentralized governance mechanisms with Web3 technologies, including smart contracts and DAOs (Decentralized Autonomous Organizations). Decentralized governance has a flat hierarchy consisting of multiple managers who may no one has to know or trust anyone else where each member in the network has a copy of the same data in the form of a distributed ledger. If the ledger has been altered or corrupted, it will be rejected by the majority of the members in the network. This would allow for the depreciation of misalignment of interest because the managers act as both the principal and the agent to avoid the possibility of misalignment regarding the same group of people [22]. Decentralized governance is popular among blockchain and DeFi projects (e.g., Uniswap, Yearn Finance, dYdX, MakerDAO, Synthetix) [3], [4]. According to DeepDAO, over four thousand DAOs were in operation in early 2022 [23]. DAOs are virtual organizations that blockchains and smart contracts can support. Members of a DAO can propose, vote, and enact changes to the virtual organization. A DAO can enable its stakeholder community to work collaboratively towards achieving shared goals. It can operate without a conventional institutional structure or a traditional decision-making hierarchy in centralized organizations where roles at the top often hold significantly more control over critical decisions than those lower in the hierarchy. The governance of a DAO can be bottom-up-based and community-driven. Its decision-making process is more inclusive and transparent compared with centralized governance. DAOs are easy to launch and customize with proper incentive structure design; DAOs can better align participant interests in a complex stakeholder environment [24].

Decentralization technology relies heavily on the blockchain, adding benefits to the organization, such as scalability, where smart contracts and digital assets support organizations that operate on the Internet and can scale globally from birth. Leveraging blockchain's tamper-proof and immutable properties provides greater transparency and responsiveness than traditional organizations. The decision-making process is publicly available knowledge and recorded on the blockchain, reducing costs, leading to more frequent voting, and enhancing data quality since the blockchain will not have duplicate, missing, or noisy data. This results in a more deterministic structure, increasing

accountability and reducing fraud, which decreases monitoring and enforcement costs [22].

E. MODEL-DRIVEN ENGINEERING OF MLOps OR DevOps

MLOps, a combination of ML (Machine Learning) and DevOps (Development Operations), encompasses best practices that integrate ML development and operations within a unified framework. It involves various components such as ML data sources and datasets, repositories of ML models and metadata, automated ML pipelines, and workflow. The application of model-driven engineering in MLOps or DevOps is still early, despite the extensive research conducted in MLOps within a centralized setting, mainly driven by the widespread adoption of cloud-based MLOps services [25].

The dynamic and ever-changing nature of MLOps poses challenges when modeling the process using model-driven engineering tools and processes. However, recent research has focused on exploring the application of model-driven engineering in centralized DevOps, and IT systems [26]. The model-driven engineering approach highlights the significance of placing requirements in the appropriate business context, identifying key processes and use cases, and ensuring effective communication with external systems. Additionally, non-functional requirements, such as performance, safety, reliability, and usability, are crucial considerations in this context. Further research delves into relevant studies, provides an overview of the model's components, introduces UML profiles specifically tailored for integration flows and blockchain deployment, and presents a design pattern for smart contracts. For example, [27] is demonstrated through three case studies, showcasing an integration flow diagram for prescription management. These examples illustrate how the model can be implemented in real-world scenarios.

III. OVERVIEW OF DECENTRALIZED ML GOVERNANCE

In this section, we propose a new framework of decentralized ML governance and describe its scope. The main objective of this endeavor is to establish a foundation of ML governance based on the principle of decentralization. In this section, we define decentralized ML governance and delineate its scope.

A. MOTIVATION

Although the prior efforts described in the previous section have taken steps to define ML governance, automate ML operations, and even explore decentralized computing infrastructure for training. They need to catch up in many aspects by not fully exploiting the potential of decentralization, particularly from an ML governance perspective. Most of the existing vision of ML governance is centralized, where a big tech company or a central entity is assumed to be responsible for taking the role of governance. Similarly, MLOps is an extension of the existing practice of DevOps to ML. The success of conventional DevOps relies on centralized tooling to offer a single toolchain and orchestration process for operation and development teams to follow across an enterprise. In the prior work, deploying ML training and inference to a decentralized computing infrastructure explores the potential of decentralization to ML. Such application occurs at the training and inference level instead of the governance level, which is the focus of this work.

With the advance of blockchains and Web3, the concept of digital ownership is expanded to a new level with decentralized, permanent data storage managed by decentralized governance mechanisms like DAOs [3], [28], [29]. This expansion opens new frontiers, such as decentralized entities' ownership of data, models, and ML code. Furthermore, with Web3 as the Internet of value, MLOps can be redefined by adding value as a new axis. The expanded definition of ML governance opens a new universe of ML value chains where ML governance manages the flows of values for ML in a complex ownership environment.

Motivated by this new vision, this paper systematically examines the landscape of decentralized ML governance and its impacts on ML systems and development. We hope that the work will pave an initial road for further research in this direction.

B. DEFINITIONS

The scope of decentralized ML governance is to support broad ML governance with decentralization using approaches like blockchains/distributed ledgers and smart contracts. The broad definition of ML governance goes well beyond security and privacy. It covers value chain management, ML finance, and community management (data engineers, DevOp engineers, model engineers, auditors, sponsors, application developers, etc.). Some announced properties of decentralized ML governance are:

- *DAO-based governance to manage the lifecycles of ML models and end-point services.*
- *ML value chain collaboration by smart contracts and DAO where blockchains can facilitate ML's value flow tracking and incentivize ML's value co-creation process.*
- *ML workflow management by a hybrid environment with both on-chain and off-chain components, which brings benefits such as transparency, accountability, and audibility.*
- *DAO-based community management of ML ecosystem participants, including decentralized identity management (e.g., DIDs).*
- *Decentralized governance of ML assets and artifacts (e.g., access control, rights management), covering data, models, and code.*

FIGURE 1 shows the architecture of decentralized ML governance and the major components.

C. DECENTRALIZED ML VALUE CHAINS AND VALUE CO-CREATION

Traditionally a single entity may perform the entire ML pipeline like data collection, model training, and model serving. The emerging trend is the involvement of multiple

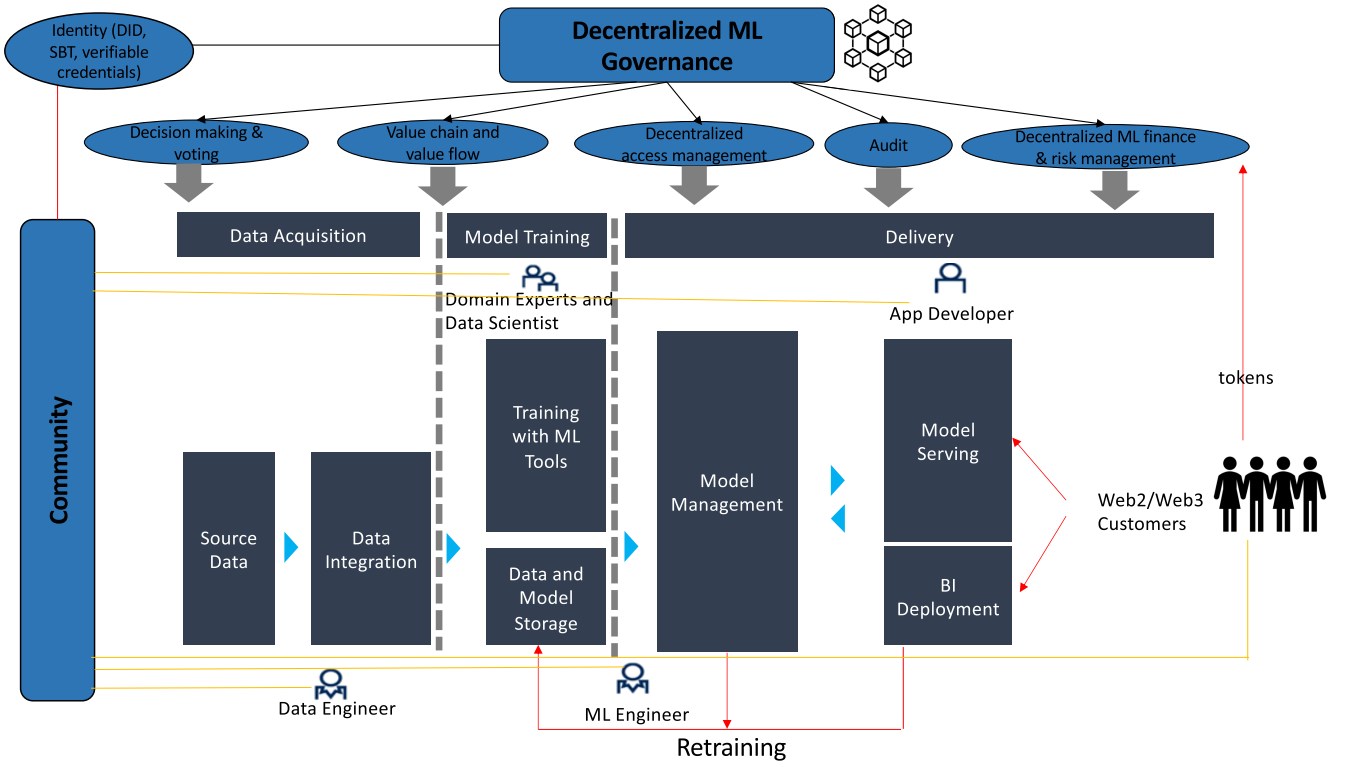

FIGURE 1. Decentralized ML Governance. MLOps define the components (gray boxes). Decentralized ML governance focuses on the blue components and the seamless integration of these components with the gray boxes.

entities in the ML pipeline where each entity is specialized in providing services of one stage, for instance, data collection and preparation, model training, or model serving. The economy is the underpinning factor that drives this trend. It is often more cost-effective and productive to have a single entity focusing on just one stage of the ML pipeline so that the services can be perfected to a highly competitive level compared with in-house approaches. This suggests that instead of viewing the ML process as a pipeline (implying that a single entity manages the process), the ML process should be treated as operations of value chains. In light of this perspective, ML governance can be considered a task of managing ML value chains where data, models, model training, fine-tuning, and model serving are goods and services. In ML value chain governance, the activities will be centered around creating or adding value to the ML artifacts (e.g., data, models, code, and services). FIGURE 2 shows the view of ML governance as a value web and its relationship with the pipeline view of the ML process.

The transformative power of casting ML governance as a value chain process is that it makes integrating ML governance with the blockchains and Web3 a natural step because blockchains are created for tracking, storing, and trading values. The concept of value co-creation originated in business management literature and practice [30], [31]. It represents a paradigm shift from considering organizations as the definers of value to a more inclusive and collaborative process involving other stakeholders and end-users. The interest in co-creation is increasingly recognized in managing and innovating value chains. From the definition of value co-creation, one can observe that the concept of co-creation is naturally aligned with the properties of blockchains and decentralized governance. As highlighted in [32] and [33], properties of blockchains, like traceability of contributions, transparency in recognizing authorship, capitalization of transactions, etc., are congruent with co-creation. In decentralized value co-creation, autonomous ML value chain stakeholders (e.g., data collectors/owners, algorithm developers, model trainers, model fine-tuners, inference service providers) can join forces and collaborate for a specific ML project or system as if they work for a single organization. Under a previously agreed upon distribution model (implemented on-chain), the revenue and income of ML services can be divided among the stakeholders. The value creation process's transparency and the stakeholders' open coordination make the approach attractive. There are many options for distributing ML values among the stakeholders, like revenue and profit sharing. Regardless of the option, the value chain process can be implemented as smart contracts, and its execution can be automated. On-chain deployment of ML value chain governance for specific ML projects can improve stakeholder and participant trust. Once the ML projects are joined, the participants are incentivized to engage and collaborate closely in the value-creation process.

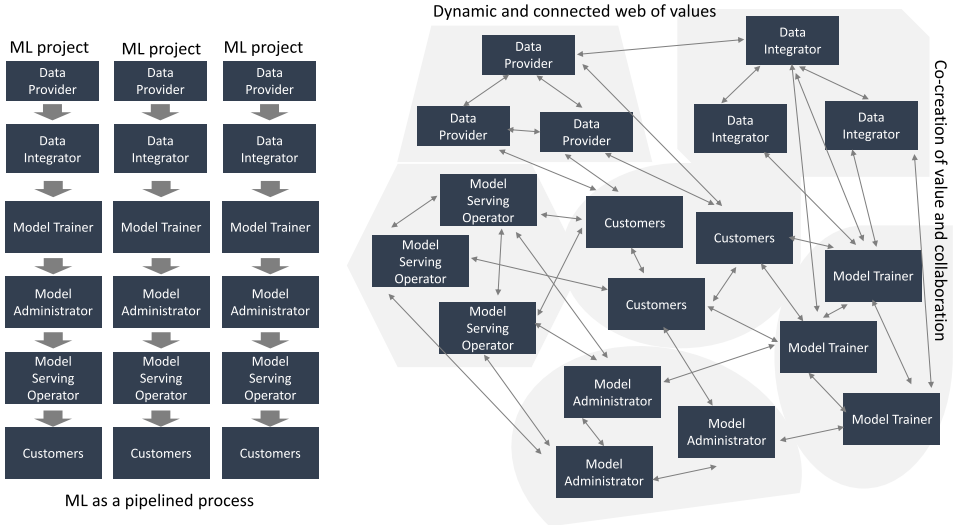

FIGURE 2. Pipeline view of ML process vs. view of ML governance as a value web with co-creation by the stakeholder community.

Decentralized ML value chains and co-creation

The ML value chain governance and the decentralized value co-creation process represent a new perspective for ML. It moves forward ML governance from pipeline process to value chain management. A decentralized value co-creation process for ML will likely spur open stakeholder collaborations and accelerate ML development and adoption. Despite the potential, decentralized ML value chain governance is relatively new. In-depth research is needed before it can be fully embraced in practice. This requires cross-disciplinary collaboration among ML, blockchains, economics, and management science fields.

D. DECENTRALIZED IDENTITY MANAGEMENT

Digital identity is a fundamental component of ML governance. With identity, it is possible to identify the stakeholders involved in the ML process and establish accountability. Identity is essential for activities such as access to ML artifacts (data, model, code), access to ML resources like computing resources for training and inference, access to ML services, participation in ML governance, and making operational decisions. ML process and governance can define multiple roles like governance, model training, operation, audit, data preparation, financial controller, risk management, etc. A person may take responsibility for multiple roles. Access control is necessary and critical to safeguarding the ML process, ensuring the integrity of ML governance, and protecting ML digital assets and artifacts. Access control can be role-based or attribute-based. For instance, an ML system can define who can update the trained model, who can authorize financial transactions, who can vote in governance decisions, and who can access training data. Compromise of identity management and access control in an ML system can result in a disastrous outcome.

IAM (Identity Access Management) is a core service for all cloud and data center providers. In the centralized MLOps model, identities are defined according to well-established standards (e.g., OAuth [34] and OpenID [35]). To support interoperability and avoid fragmentation, federated identity management [36] is developed to enable users to access the resources and services of multiple organizations using a single set of credentials. A benefit of federated identity is that it supports linking a user's identity across multiple separate identity management systems.

Another digital identity paradigm, self-sovereign identities (SSI) [37], has emerged recently. SSI is more decentralized and based on technology such as blockchains. It puts end-users entirely in control and allows different service providers to share identity verification attestations. Compared with the centralized and federated identity models, SSI's locus of control is with the issuers and verifiers in the system. In the decentralized SSI models, the control shifts to the individual identity owner, who can now interact as a full participant with everyone else in a decentralized environment. A related effort in this direction is the W3C initiative on standardizing DIDs (Decentralized Identifiers) [38]. According to W3C, a DID is a digital identifier not needing to be leased. Its creation and use do not rely on a central authority to manage it. DIDs are helpful for any application that benefits from self-administered, cryptographically verifiable identifiers such as decentralized, verifiable credentials [39] to identify people, organizations, and things to achieve desired security and privacy-protection guarantees. W3C DIDs and verifiable credentials can offer standard-based solutions to support identity guarantees in decentralized ML governance. DIDs are decentralized and self-managed, matching the decentralized governance model for ML. Meanwhile, privacy can be fully respected with techniques such as zero-knowledge proof and verification of claims [40]. For instance, stakeholders

can claim skill levels, experiences, and ownership without revealing sensitive identity data. The others in the system can verify the claims. More recently, a concept called soul-bound tokens (SBT) [41] was proposed to achieve the vision of a decentralized society. Soul-bound tokens are publicly visible and non-transferable tokens. They are defined through social coordination and certified by other related ''souls''. For instance, other ML specialists or users interacting with it in an ML community can approve a soul-bound token. The certification process is decentralized and community-based. It is not required that a soul must be a legal name or one soul per person. Whether soul-bound tokens can be tied with ML models remains an interesting question. The current definition of soul-bound tokens only partially recognizes such a scenario.

Decentralized identity

Decentralized ML governance requires a decentralized identity as part of its foundations. Concepts like SSI, SBT, and W3C DIDs/verifiable credentials provide great opportunities to achieve this goal. However, challenges remain, such as integrating DIDs/verifiable credentials with ML governance to achieve security and privacy-protection guarantees, supporting interoperability between decentralized and centralized identities in an ML system, and the meaning of soul-bound tokens for ML systems.

E. DECENTRALIZED OWNERSHIP AND DECENTRALIZED RIGHTS MANAGEMENT

ML systems include artifacts such as data (training and testing), models, and code. A plural of rights can be defined over these ML assets, such as ownership, suitable to use, right to develop derived work, and right to upgrade or modify. For instance, the owner of a dataset can license the dataset to model trainers to include it in a model training task. Holding certain rights will allow the stakeholders to perform specific actions that would otherwise be prohibited, like creating derived work, hosting an ML model as a service, and using a dataset for training. A qualified entity can grant other ML participants rights to the ML artifacts. For example, specific licenses can be issued to the users or participants of an ML system to allow them to train ML models based on a protected dataset. Licenses can have different types like permanent, renewable, term-based, etc. Further, transferring or leasing rights of the ML artifacts from one entity to another is plausible. For example, the owner of an ML model could lease the model to another entity over some time (agreed upon in the lease term) so that the lessee is granted the right to obtain the economic benefit from using the ML model. However, the model still belongs to the original owner.

With the decentralization of ML governance, the landscape of digital rights to ML artifacts becomes more complex:

- The entity that holds certain rights to ML artifacts can be an online community or a virtual organization like a DAO.
- Digital rights, ownership, and license management can be decentralized. For example, a license to ML artifacts can be transferred from one virtual organization to another; a community of online participants can grant the rights to use specific ML assets to other entities.
- In a decentralized environment, the identities of the participants and stakeholders can be based on SSI or decentralized (DID or SBT based).

Many research questions and challenges arise from the decentralized governance of digital rights and ownership of ML assets and artifacts, for instance, how to ensure data integrity, data confidentiality, and rights protection when a community of stakeholders or a virtual organization like a DAO owns ML assets. How to manage the licensing process when the issuer is a decentralized virtual organization? How do we audit if a community owns the ML artifacts? How to resolve disputes when there is a disagreement about ownership or rights between two virtual organizations?

Prior work exists to leverage blockchains and distributed ledger technology for digital rights management [42], [43], [44]. For instance, smart contracts can be leveraged for managing copyright transactions and issuing licenses automatically, eliminating the need for centralized entities to verify identities and issue licenses. Most of the efforts focus on the traditional use cases of IP protection like copyrights and take advantage of blockchain characteristics such as immutability and auditability to track IP ownership and licenses. The scope of these efforts is quite limited compared with what is needed for decentralized ML governance. For instance, a challenge of ML governance is how to guarantee the confidentiality of ML assets and protect the owners' interests in a decentralized manner when they are used for training or inference. A related area to decentralized ML governance's challenges is decentralized access control management. Recently, blockchain-based access control has been intensively studied (e.g., [45], [46]). Although these research efforts do not target the use cases of rights management for decentralized ML governance, they could provide specific reusable components or technology tools to develop a solution applicable to decentralized ML governance.

FIGURE 3 demonstrates a possible scenario where a community of stakeholders owns ML assets. The assets at rest are protected with a suitable encryption scheme, for instance, threshold cryptographic system [47], [48]. The encrypted ML assets can be stored in decentralized storage like IPFS or other Web3 storage systems. When the assets are needed for training or serving, a right holder (a participant who has the right to train a model using the protected data or a participant who has the right to use the protected model) can present evidence of its identity and right to the stakeholder community who jointly owns the keys for decryption. After successful verification, the key owners can release sub-keys (key shares) to the right holder. Then, the right holder can

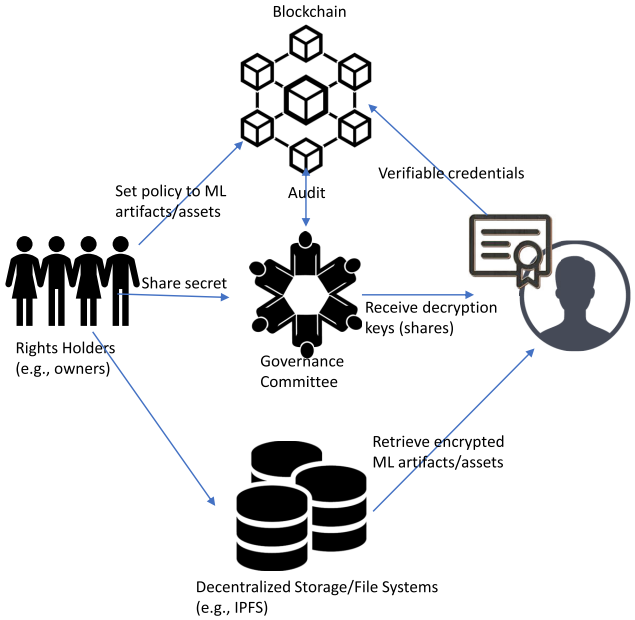

FIGURE 3. Decentralized access management of ML assets/artifacts.

assemble the decryption key to decrypt the ML assets. It is worth mentioning that the right holder, in this case, can be a human being, a machine, a computer cluster, or a virtual organization represented by its digital identity described earlier.

It is plausible to protect ML data, models, and even code when used (e.g., training and inference). Several technology frontiers are under active research and development to provide such solutions. Homomorphic ML is one area where ML tasks can be performed over encrypted data [49], [50]. Despite heavy research in homomorphic ML performance improvement, existing approaches still suffer from low performance. In addition, it is mainly suited for ML inference instead of training [51] because of the computation cost. To protect the confidentiality of the training data, federated ML is a promising direction to explore [52]. Researchers have started integrating federated ML with smart contracts and blockchains [53], [54]. Another direction is to leverage hardware with specific security features, such as Trusted Execution Environments (TEEs), for ML (e.g., [55]). TEE-based ML can deliver better performance compared to other options like homomorphic ML. However, TEE-based approaches face challenges, such as a lack of vendor-agnostic standards in TEE implementation, low performance compared with non-TEE-based ML, and vulnerabilities like side-channel exploits and other exposed attack surfaces (e.g., [56], [57], [58], [59]). Verification of the ML process can be either centralized or decentralized. It is preferred to support decentralized verification schemes for decentralized governance. Supporting decentralized ML governance may require capabilities of public verification, for instance, community-based proof of ownership of certain rights to the ML assets. A challenge of decentralized verification is privacy, which could be solved with zero-knowledge-based protocols. Other related research topics include watermarking of ML models [60], [61], verification of derived work in ML, for instance, proof of an ML model trained based on a given dataset. Solving research challenges in these topics may involve the development of new zero-knowledge-based approaches [62], or ML-oriented verifiable computations where computation and transformation applied to ML models can be publicly verified.

Decentralized ownership and rights management

Managing the rights of ML assets in decentralized ML governance is a rich ground for research. ML governance involves digital rights management of ML assets (e.g., data, models, code). Cryptographic primitives are needed to safeguard the integrity and confidentiality of the ML assets not only when the assets are at rest but also when the assets are in use (e.g., training, inference). Using blockchain-based technology, various rights to ML assets can be managed on-chain by a community of stakeholders. Furthermore, decentralized ML governance can benefit from the technologies that allow public verification of ML processes, such as verification that an ML process upholds pre-agreed upon license terms, verification of the integrity of the ML process and verification of ML asset ownership.

F. OPPORTUNITIES OF DECENTRALIZED RISK MANAGEMENT

ML process, by nature, involves risks in its whole life cycle. The risks can be security and/or privacy-related, such as disclosure of private training data by model inversion attacks, theft of copyrighted ML models, and unreliable ML predictions due to the poor robustness of the ML models. The risks of ML systems can be societal and financial. For instance, the fairness of ML models would affect society's social justice and well-being. There is a need for service level agreements (SLA) for ML services and systems. The existing ML service paradigm is economically biased toward the service providers instead of the end-users because the service providers are not held accountable financially for the potential damage that the provided services can incur to the users. This deficiency must be remedied. Otherwise, it may hinder society's wider adoption of ML and undermine the trust between the ML service providers and the end-users. For example, ML can be applied to automate financial transactions in DeFi, facilitate business processing, act as Oracle sources [63] for dApps, and control cyber-physical systems. When an ML system fails to deliver its services at the promised level of quality, like incorrect predictions, the ML service providers should be financially accountable for the damage or loss incurred to the end-users. For example, an ML-based Oracle source may provide an incorrect data feed to DeFi applications. One can quickly develop similar use cases of ML where

the end-users desire some QoS guarantee and SLA. Under the broad umbrella of ML governance, solutions should be provided to satisfy the users' needs. Blockchains generally have three risk management approaches reputation-based, staking-based, and insurance-based. Each approach has its pros and cons. In a fully decentralized and permissionless environment where ML service providers (e.g., data sources, model providers, inference services) are anonymous, staking is a suitable approach. Many dApps apply to stake for managing trust and risks. However, staking has its downsides, for instance, high cost to the service providers and efficiency issues due to the lack of financial assets during staking. Decentralized insurance [64] and risk management are attractive because these approaches can lower the cost for the stakeholders. In addition, decentralized ML risk management enriches the scope of ML governance by offering new research opportunities like decentralized insurance for ML. Rigorous risk modeling and assessment of ML system risks based on solid theoretical foundations in ML likely hold the key to the success of decentralized ML insurance. Modeling and pricing financial and economic risks involving ML systems are relatively new research topics.

Opportunities for decentralized risk management

ML governance should include managing ML-related risks, including financial and economic risks. Pricing and modeling financial and economic risks of ML applications are essential for the societal-level success of ML systems because they provide economic fairness and assurance to the end-users, particularly for applications where ML services will be integrated for automated operations. Decentralized ML risk management has certain advantages, such as risk pooling, leveraging community wisdom for risk assessment, and cost reduction by technology, such as decentralized insurance. Problems like pricing, risk modeling, decentralized insurance, and risk management for ML systems and services remain open research questions.

G. DECENTRALIZED DECISION-MAKING PROCESS

Governance involves making decisions. In ML governance, one can provide numerous scenarios where decisions are needed to manage and control ML processes, such as a decision to expand training dataset, a decision to include a specific dataset into model training, a decision to adopt a particular design of an ML model, a decision to support particular ML use case, a decision to integrate ML models for an application, a decision to license a model to other users or virtual organizations, a decision to reward the contributors to an ML model. Decentralized governance means a decentralized decision-making process.

A specific example of applying decentralized governance for ML is OpenAI's project on democratic inputs for LLM (large language models). LLMs like ChatGPT are fine-tuned based on human feedback using a technique called Reinforcement Learning from Human Feedback (RLHF). This process ensures that responses from the LLMs are aligned with so-called human values. However, designing an open, fair, transparent, and accountable alignment process poses significant challenges, as only some sources of preferences can fully represent the diverse values of all stakeholders, including the organization training the model, customers, end-users, and the broader population [65]. For instance, the GPT models were fine-tuned using RLHF based on feedback from a small set of human trainers recruited by OpenAI, mainly from South Asia.

Since human feedback for fine-tuning plays a crucial role in determining the alignment of the LLMs and human values are diverse and global, OpenAI recognizes this challenge and proposes research to improve the process of tuning LLMs so that they are capable of being conditioned on the preferences of specific groups or easily fine-tuned to represent different values. The project aims to ensure that all human groups are represented and protected from potentially harmful processes, emphasizing the need for aligning LLM feedback governance with the collective good to reduce potential risks to human society. One approach to achieve such alignments is to involve the public in shaping the LLMs. This could leverage decentralized autonomous organizations (DAOs) characteristics to gather collective inputs from the public and implement values trade-offs, which could foster a more inclusive and responsible future in creating LLMs [66].In decentralized ML governance, the decision-making process is decentralized as a community-led effort with no central authority. The process can occur in the blockchain space involving either on-chain decision-making or a hybrid approach of off-chain decision-making (e.g., off-chain voting) and on-chain finalization of the decisions. Without central leadership, decentralized ML governance can be realized as virtual organizations such as DAOs [3]. In this case, decisions are made from the bottom up and governed by the community participants in the ML project. When smart contracts are employed, decisions can be supported by different voting strategies and rules, implemented based on weighted voting, delegate voting, ranked-choice voting, etc. Decentralization governance hypothesizes that community-based governance can result in better decisions if designed properly than centralized governance. Whether decentralized ML governance can lead to better decisions remains to be tested. However, the bottom-up decision process has certain advantages for aligning the interests in a multi-stakeholder environment like ML systems. Besides the research questions above, decentralized ML governance faces the challenges such as privacy protection (privacy-preserving voting [67]), defense against attacks and manipulation of the decision-making/voting process (e.g., [68],

mitigation of governance risks, fairness in the governance process.

Decentralized decision-making process

ML governance must make decisions to manage the ML process. Decentralized ML governance offers opportunities for managing ML processes and systems based on community decisions and choices. Although it remains an open hypothesis whether decentralized ML governance can eventually lead to better decisions for ML governance, the potential benefits of decentralized governance and community-driven decision-making are well recognized. Many open research questions and challenges exist, such as privacy, tooling for DAO-based decision-making, and mitigating governance risks. These questions require cross-disciplinary collaboration among computer science, management, and behavioral sciences.

H. DECENTRALIZED ML FINANCE

ML finance is essential to ML governance. Trained ML models are at the center of ML systems because of their potential to support a diverse and large number of impactful applications (e.g., GPT- 3 [69], stable diffusion model [70], M6 model by Damo Academy [71]). Over the years, these general-purpose ML models have grown, becoming even more extensive at a pace far exceeding the growth of hardware speed limited by Moore's Law. Consequently, it becomes increasingly expensive to train and own these models. The Allen Institute for AI puts the average cost to train an ML model at $1 / 1000 parameters. As the parameters increase to the range of trillions, so does the cost to train these models. According to estimate [72], a billion parameter model could have a price tag of about $1M. This means that anytime soon, only very few large tech companies can afford the cost of training such large ML models. This means that the ML process has a looming financing problem, which will worsen. To solve the ML financing challenge, decentralized ML governance can benefit from the rich space of decentralized finance [73]. Various purpose-built DAOs can be set up to finance ML systems and processes, such as a donation DAO for ML projects, a consortium DAO for a specific ML system, a crowd-funding DAO for specific ML services, and a revenue-sharing DAO for ML systems. An ML DAO can be created to raise capital to fund ML projects for public goods. Sooner or later, we will see a large web of connected ML DAOs to finance ML projects and services. This will open almost unlimited opportunities in research and practice in ML finance. For example, given a finite amount of resources and a complex environment of ML projects and systems, where should the resources be spent to get the best bang for the investment? In the case of financing ML projects and systems for social goods, how to measure the social impacts of ML services? How to quantify the return on investment when ML governance is applied to improve social goods? How to allocate financial resources optimally in the context of a web of ML projects to maximize the returns?

Decentralized ML finance

Decentralized ML governance and decentralized finance open many opportunities in research and practice. With the rising cost of owning general-purpose ML models, the advance of ML certainly depends on the success of ML finance. There are many research questions, such as optimizing the returns from the investment to a portfolio of ML projects. How to govern the decentralized framework of ML finance.

IV. CHALLENGES FOR DECENTRALIZED ML GOVERNANCE

A. DAO GOVERNANCE CHALLENGES

Despite the advantages of decentralized governance, DAOs also have many limitations and potential disadvantages. Clearly defining the roles, responsibilities, and incentives for the stakeholders and contributors, managing large stakeholder communities using off-chain communication channels, and monitoring the community's needs are often resource and labor-intensive tasks. Due to the cost of on-chain voting, it is common for DAOs to delegate governance authority to a small committee where the committee has significant power over the DAO members. Many studies of popular DeFi projects have observed actual centralization or plutocracy of the governance mechanisms (e.g., [74], [75], [76]. For many projects, community engagement is low. Most community stakeholders do not actively participate in governance, either abstaining completely or ceding their power to the protocol development team or so-called ''protocol politicians''.

Other challenges of blockchain-based governance besides voter turnout include voter fatigue, manipulation of the voting process, voter bribery, and other attacks on DAO-based decision-making [77]. For instance, in optimistic voting, proposals are set to be adopted by default unless a quorum of voters objects. When votes are weighted, governance may be dominated by few participants with more resources than the others in the community [74], [78]. In the case of hybrid governance combining off-chain and on-chain voting, it often takes a long delay for the off-chain voting decisions or proposals to be reflected. Further, free riders can be found in DAO-based communities.

B. SECURITY CHALLENGES

Decentralized governance, implemented through smart contract code and/or off-chain software working in conjunction with on-chain code, introduces potential security

vulnerabilities that malicious actors can exploit. While blockchain-based federated learning offers decentralized and privacy-enhancing data processing capabilities, it is not immune to security risks. Prior research has addressed security concerns, including backdoor and eclipse attacks. Proposed approaches aim to accelerate the propagation of backdoors, reduce attack costs, and identify novel hybrid vulnerabilities for spreading backdoor attacks among swarm learning nodes [19], [79].

Various techniques have been explored to mitigate security vulnerability attacks from users and protect local model updates from the server. Secure multiparty computation allows multiple parties to perform computations without sharing their data, preventing data poisoning and model inversion attacks. Differential privacy adds noise to data, safeguarding sensitive information while maintaining data utility and preventing membership inference attacks [80]. Also, multi-tentacle federated learning (MTFL) and its application in the software-defined industrial Internet of Things (SD-IIoT)guarantee the reliability of training data by clustering participants with similar learning tasks into tentacle groups. This grouping strategy is crucial in detecting adaptive poisoning attacks using the tentacle distribution-based efficient poisoning attack detection (TD-EPAD) algorithm [81]. Homomorphic encryption enables computation on encrypted data, preventing unauthorized access to sensitive information. Combining homomorphic encryption and federated learning reduces the risk of data integrity and security breaches by analyzing data in different locations based on raw data exchange and model updates [82].

Despite advancements in smart contract audits, automated vulnerability detection, and formal verification of smart contract properties(e.g., [83], [84], [85], [86], it is impossible to guarantee the absence of vulnerabilities in governance software. Successful exploits of decentralized autonomous organizations (DAOs) can have severe consequences, impacting the stakeholder community's reputation, trust, and financial interests.

C. INTEROPERABILITY AND INTEGRATION CHALLENGES

The scope of decentralized ML governance includes many technological areas, from decentralized identity, decentralized access control and rights management of ML assets, and verifiable ML to the ML value chain and value network, decentralized ML finance, and risk management. Most of the areas can be studied separately. Specific technology and standards could be developed to solve a sub-problem within each area, for instance, the standard for decentralized identity or decentralized management of digital rights. Integrating each area's results and research outcomes into a complete solution for ML governance with interoperability is a challenge. Further, it is certain that decentralized ML governance needs to interact with the traditional computing and service environment, like taking advantage of the cloud-based infrastructures for training. Interoperability between blockchain-based governance and the traditional non-blockchain-based environment is unavoidable for decentralized ML governance.

D. PRIVACY CHALLENGES

Decentralized governance in machine learning (ML) must address privacy challenges, including identity management, ownership and rights management, access control, voting, audibility, and maintaining privacy and confidentiality throughout ML processes. Research questions regarding confidentiality in decentralized ML governance still require answers. While privacy audibility is a priority, accountability and trustworthiness must also be maintained. Many decentralized autonomous organizations (DAOs) operate anonymously, and strong privacy can be achieved through zero-knowledge-based protocols for fully anonymous voting. While privacy protection allows stakeholders to participate without revealing their identities, it can be a double-edged sword for the long-term well-being of decentralized governance, as bad actors may exploit strong privacy for unethical or unlawful actions. Proper management and design of governance mechanisms based on privacy-preserving technologies are necessary to prevent a lack of accountability and hinder the acceptance of decentralized governance.

Decentralized ML governance must also integrate with traditional computing and service environments, using cloud-based infrastructures for ML model training. This requires interoperability between blockchain-based governance and non-blockchain-based environments, essential for effective decentralized ML governance. In IoT devices, the BDEV-CAML technique addresses challenges related to faults and reliability. It combines blockchain technology, IoT, and machine learning models to enhance trustworthiness, effectiveness, and security in IoT networks. By leveraging these technologies, BDEV-CAML provides a robust and secure approach to IoT fault detection, addressing data integrity and privacy concerns in sectors such as healthcare [87]. Although significant research has been conducted on privacy in permissioned blockchains, and efforts have been made to tackle privacy challenges in decentralized machine learning governance through the use of differentially private techniques and error-based aggregation rules to enhance privacy, prevent attacks, and improve resilience [88], there remain numerous open questions that necessitate further exploration. These questions pertain to various aspects of privacy in decentralized machine learning, such as confidentiality, privacy audibility, and the balance between privacy and accountability. Continued research and investigation are crucial for advancing our understanding of these open questions and developing comprehensive solutions to ensure robust, privacy-preserving, decentralized machine learning governance.

E. LEGAL AND REGULATORY CHALLENGES

Decentralized governance based on DAOs faces legal and regulatory challenges, such as the uncertainty of legal status. Without legal status certainty, DAOs are not protected as legal entities. The participants, though located around the globe, are legally liable for their actions of the DAO. This means that DAOs cannot use legal protections such as limited liability and take advantage of economic benefits like tax credits typically given to legal organizations. In addition, data privacy-related regulations like GDPR [89] and CCPA (California Consumer Privacy Act) [90] could pose compliance challenges for decentralized ML governance where management of the ML assets are decentralized. There is also a trend of the growing number of legislation explicitly targeting ML and AI [91].

For centralized and decentralized ML governance, compliance and legal audit would be significant challenges from both technical and legal aspects. For instance, a decentralized organization manages compliance requirements, responds to requests from regulatory agencies, and operates to meet all consumer demands on time. If access control to ML assets and artifacts is decentralized using cryptographic primitives, how can the decentralized governance body act according to the legal requirements? In the case of ML value chain/web involving multiple stakeholders and distributed resources, providing evidence of compliance for audit remains challenging.

Blockchain technology is a fundamental component of decentralized autonomous organizations (DAOs). On a positive note, some countries and states accepted blockchain as evidence in court. Several states have introduced legislation and regulations to regulate blockchain and cryptocurrency in the United States, including New York's BitLicense. Additionally, some states have passed or introduced legislation specifically regulating the admissibility of blockchain evidence in court, such as Illinois, Vermont, Virginia, Washington, Arizona, New York, and Ohio.

Many countries, including China, Azerbaijan, the United Kingdom, the United States, and Italy, have made legal changes to accept blockchain evidence. China's Internet courts have even taken blockchain evidence in two separate cases, setting a precedent for other legal systems. Despite these developments, most courts must implement legal changes to accept blockchain evidence. As the industry grows and reliance on distributed ledger technology (DLT) increases, legal certainty must be established in every jurisdiction seeking to keep pace with this technological advancement [92].

V. FUTURE WORK

This paper presents a framework for Decentralized Machine Learning (DML) Governance, providing insights into the opportunities and challenges of this innovative framework. To further advance decentralized ML governance, conducting in-depth examinations and studying specific use cases is crucial. By analyzing detailed use cases, researchers can better understand the framework and identify potential implementation challenges, leading to the proposal of practical and viable solutions.

A practical approach to enhancing our understanding involves focusing on case studies and examples across different contexts. Such studies will provide valuable insights into the applicability of decentralized ML governance in various domains and industries. Additionally, conducting thorough analyses of challenges faced in real-world scenarios, such as model security, data privacy, and adversarial attacks, will be essential for building robust and trustworthy decentralized governance for ML systems.

Future work directions include delving into practical implementations and case studies, such as DAO tooling, governance contract design patterns, and demonstration using case studies, as iteratively fine-tuning governance mechanisms, addressing ethical and regulatory considerations, exploring interoperability and standardization, including a model-driven extension to decentralized ML governance. By taking these future work steps, one can unlock the full potential of decentralized ML governance and usher in a new era of trustworthy, transparent, and responsible ML processes. Additionally, leveraging insights gained from decentralized governance research and bridging knowledge gaps can strengthen the connection between ML governance and DAO research. This integrated effort will drive the advancement of both fields and accelerate the adoption of decentralized governance in diverse applications and industries.

VI. CONCLUSION

Machine learning in computer systems introduces many benefits but also raises risks to society, indicating the importance of introducing the concept of governance based on the principle of decentralization. The scope of decentralized ML governance is to support broad ML governance with decentralization by taking advantage of approaches like blockchains, distributed ledgers, and smart contracts. The definition of ML governance goes well beyond security and privacy, which covers value chain management, ML finance, community, and management. In this paper, we study the details of the decentralized ML governance framework in-depth and provide a comprehensive view of its components, research opportunities, challenges, and open problems. While not providing a prototype, this paper can serve as a roadmap to facilitate the research and development of decentralized ML governance.

REFERENCES

[1] A. W. Senior, R. Evans, J. M. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Zídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, and D. Hassabis, "Improved protein structure prediction using potentials from deep learning," *Nature*, vol. 577, no. 7792, pp. 706–710, 2020.

[2] V. Chandrasekaran, H. Jia, A. Thudi, A. Travers, M. Yaghini, and N. Papernot, "SoK: Machine learning governance," 2021, *arXiv:2109.10870*.

[3] S. Hassan and P. De Filippi, "Decentralized autonomous organization," *Internet Policy Rev.*, vol. 10, no. 2, pp. 1–10, Apr. 2021. [Online]. Available: http://hdl.handle.net/10419/235960

[4] R. V. Pelt, S. Jansen, D. Baars, and S. Overbeek, "Defining blockchain governance: A framework for analysis and comparison," *Inf. Syst. Manage.*, vol. 38, no. 1, pp. 21–41, Jan. 2021, doi: 10.1080/10580530.2020.1720046.

[5] V. Sridhar, S. Subramanian, D. Arteaga, S. Sundararaman, D. Roselli, and N. Talagala, "Model governance: Reducing the anarchy of production ML," in *Proc. USENIX Annu. Tech. Conf. (USENIX ATC)*. Boston, MA, USA: USENIX Association, Jul. 2018, pp. 351–358. [Online]. Available: https://www.usenix.org/conference/atc18/presentation/sridhar

[6] L. Yue, H. Junqin, Q. Shengzhi, and W. Ruijin, "Big data model of security sharing based on blockchain," in *Proc. 3rd Int. Conf. Big Data Comput. Commun. (BIGCOM)*, Aug. 2017, pp. 117–121.

[7] W. Xiong and L. Xiong, "Smart contract based data trading mode using blockchain and machine learning," *IEEE Access*, vol. 7, pp. 102331–102344, 2019.

[8] T. Hewa, M. Ylianttila, and M. Liyanage, "Survey on blockchain based smart contracts: Applications, opportunities and challenges," *J. Netw. Comput. Appl.*, vol. 177, Mar. 2021, Art. no. 102857.

[9] L. Cardoso Silva, F. Rezende Zagatti, B. Silva Sette, L. Nildaimon dos Santos Silva, D. Lucrédio, D. Furtado Silva, and H. de Medeiros Caseli, "Benchmarking machine learning solutions in production," in *Proc. 19th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2020, pp. 626–633.

[10] Digital Journal. *Mlops Market Size*. Accessed: Oct. 23, 2022. [Online]. Available: https://www.digitaljournal.com/pr/mlops-market-size-2022-global-growth-opportunities-cagr-value-swot-analysis-share-and-forecast-to-2028

[11] R. Subramanya, S. Sierla, and V. Vyatkin, "From DevOps to MLOps: Overview and application to electricity market forecasting," *Appl. Sci.*, vol. 12, no. 19, p. 9851, Sep. 2022.

[12] S. Peng, Y. Yang, M. Mao, and D.-S. Park, "Centralized machine learning versus federated averaging: A comparison using MNIST dataset," *KSII Trans. Internet Inf. Syst.*, vol. 16, no. 2, pp. 742–756, 2022.

[13] I. Hegedűs, G. Danner, and M. Jelasity, "Gossip learning as a decentralized alternative to federated learning," in *Proc. IFIP Int. Conf. Distrib. Appl. Interoperable Syst.* Cham, Switzerland: Springer, 2019, pp. 74–90.

[14] H. Zhu, H. Zhang, and Y. Jin, "From federated learning to federated neural architecture search: A survey," *Complex Intell. Syst.*, vol. 7, no. 2, pp. 639–657, Apr. 2021.

[15] A. P. Balcerzak, E. Nica, E. Rogalska, M. Poliak, and O.-M. Sabie, "Blockchain technology and smart contracts in decentralized governance systems," *Administ. Sci.*, vol. 12, no. 3, p. 96, Aug. 2022.

[16] F. Zerka, V. Urovi, A. Vaidyanathan, S. Barakat, R. T. H. Leijenaar, S. Walsh, H. Gabrani-Juma, B. Miraglio, H. C. Woodruff, M. Dumontier, and P. Lambin, "Blockchain for privacy preserving and trustworthy distributed machine learning in multicentric medical imaging (C-DistriM)," *IEEE Access*, vol. 8, pp. 183939–183951, 2020.

[17] B. Camajori Tedeschini, S. Savazzi, R. Stoklasa, L. Barbieri, I. Stathopoulos, M. Nicoli, and L. Serio, "Decentralized federated learning for healthcare networks: A case study on tumor segmentation," *IEEE Access*, vol. 10, pp. 8693–8708, 2022.

[18] Z. Wang and Q. Hu, "Blockchain-based federated learning: A comprehensive survey," 2021, *arXiv:2110.02182*.

[19] Z. Yang, G. Li, J. Wu, and W. Yang, "Propagable backdoors over blockchain-based federated learning via sample-specific eclipse," in *Proc. IEEE Global Commun. Conf.*, Jun. 2022, pp. 2579–2584.

[20] X. Yang, S. Wang, F. Li, Y. Zhang, W. Yan, F. Gai, B. Yu, L. Feng, Q. Gao, and Y. Li, "Ubiquitous verification in centralized ledger database," in *Proc. IEEE 38th Int. Conf. Data Eng. (ICDE)*, May 2022, pp. 1808–1821.

[21] X. Yang, Y. Zhang, S. Wang, B. Yu, F. Li, Y. Li, and W. Yan, "LedgerDB: A centralized ledger database for universal audit and verification," *Proc. VLDB Endowment*, vol. 13, no. 12, pp. 3138–3151, Aug. 2020.

[22] A. Wright. (Jun. 30, 2021). *The Rise of Decentralized Autonomous Organizations: Opportunities and Challenges*. Stanford Journal of Blockchain Law & Policy. [Online]. Available: https://stanford-jblp.pubpub.org/pub/rise-of-daos.

[23] B. Quarmby. (Dec. 31, 2021). *Dao Treasuries Surged 40x in 2021: Deepdao*. [Online]. Available: https://cointelegraph.com/news/dao-treasuries-surged-40x-in-2021-deepdao

[24] Y. El Faqir, J. Arroyo, and S. Hassan, "An overview of decentralized autonomous organizations on the blockchain," in *Proc. 16th Int. Symp. Open Collaboration*, 2020, pp. 1–8.

[25] A. Colantoni, L. Berardinelli, and M. Wimmer, "DevOpsML: Towards modeling DevOps processes and platforms," in *Proc. 23rd ACM/IEEE Int. Conf. Model Driven Eng. Lang. Syst., Companion Proc.*, Oct. 2020, pp. 1–5, doi: 10.1145/3417990.3420203.

[26] H. da Gião, "A model-driven approach for DevOps," in *Proc. IEEE Symp. Vis. Lang. Human-Centric Comput. (VL/HCC)*. Los Alamitos, CA, USA: IEEE Computer Society, Sep. 2022, pp. 1–3, doi: 10.1109/vl/hcc53370.2022.9833125.

[27] T. Górski, "The 1+5 architectural views model in designing blockchain and IT system integration solutions," *Symmetry*, vol. 13, no. 11, p. 2000, Oct. 2021.

[28] C. Hackly. (Jun. 1, 2021). *What are Daos and why you Should Pay Attention*. [Online]. Available: https://www.forbes.com/sites/cathyhackl/2021/06/01/what-are-daos-and-why-you-should-pay-attention/?sh=1f879a467305

[29] A. Zwitter and J. Hazenberg, "Decentralized network governance: Blockchain technology and the future of regulation," *Frontiers Blockchain*, vol. 3, p. 12, Mar. 2020.

[30] C. Grönroos and A. Ravald, "Service as business logic: Implications for value creation and marketing," *J. Service Manage.*, vol. 22, no. 1, pp. 5–22, Mar. 2011.

[31] S. L. Vargo and R. F. Lusch, "Institutions and axioms: An extension and update of service-dominant logic," *J. Acad. Marketing Sci.*, vol. 44, no. 1, pp. 5–23, Jan. 2016.

[32] M. Mačiulienė and A. Skaržauskienė, "Conceptualizing blockchain-based value co-creation: A service science perspective," *Syst. Res. Behav. Sci.*, vol. 38, no. 3, pp. 330–341, May 2021, doi: 10.1002/sres.2786.

[33] E. Seulliet. (2016). *Open Innovation, co-Creation: Why Blockchain is a Small Revolution*. [Online]. Available: https://medium.com/@ericseulliet/open-innovation-co-creation-whyblockchain-is-asmall-revolution-73e7d0b480d5

[34] D. Hardt, *The Oauth 2.0 Authorization Framework, Internet Requests for Comments*, RFC Editor, document RFC 6749, Oct. 2012. [Online]. Available: http://www.rfc-editor.org/rfc/rfc6749.txt. http://www.rfc-editor.org/rfc/rfc6749.txt

[35] N. Sakimura, J. Bradley, and M. Jones. (2014). *OpenID Connect Dynamic Client Registration 1.0 Incorporating Errata Set 1. OpenID Foundation*. [Online]. Available: http://openid.net/specs/openid-connect-registration-1_0.html

[36] D. W. Chadwick, *Federated Identity Management*. Berlin, Germany: Springer, 2009, pp. 96–120, doi: 10.1007/978-3-642-03829-7_3.

[37] R. Soltani, U. T. Nguyen, and A. An, "A survey of self-sovereign identity ecosystem," *Secur. Commun. Netw.*, vol. 2021, pp. 1–26, Jul. 2021, doi: 10.1155/2021/8873429.

[38] W3C. (Nov. 8, 2020). *Decentralized Identifiers (DIDS) v1.0. Core Architecture, Data Model, and Representations*. [Online]. Available: https://www.w3.org/TR/2020/WD-did-core-20201108/

[39] M. Sporny, G. Noble, D. Longley, D. C. Burnett, B. Zundel, and K. D. Hartog. (2019). *Verifiable Credentials Data Model 1.0*. [Online]. Available: https://www.w3.org/TR/verifiable-claims-data-model/

[40] U. Feige, A. Fiat, and A. Shamir, "Zero-knowledge proofs of identity," *J. Cryptol.*, vol. 1, no. 2, pp. 77–94, Jun. 1988.

[41] M. Zoltu. (May 5, 2022). *Eip-5114: Soulbound Badge*. [Online]. Available: https://ethereum-magicians.org/t/eip-5114-soulbound-token/9417

[42] Z. Zhang and L. Zhao, "A design of digital rights management mechanism based on blockchain technology," in *Proc. Int. Conf. Blockchain*, 2018, pp. 32–46.

[43] A. Garba, A. D. Dwivedi, M. Kamal, G. Srivastava, M. Tariq, M. A. Hasan, and Z. Chen, "A digital rights management system based on a scalable blockchain," *Peer Peer Netw. Appl.*, vol. 14, no. 5, pp. 2665–2680, Sep. 2021.

[44] M. Holland, J. Stjepandic, and C. Nigischer, "Intellectual property protection of 3D print supply chain with blockchain technology," in *Proc. IEEE Int. Conf. Eng., Technol. Innov. (ICE/ITMC)*, Aug. 2018, pp. 1–8.

[45] D. Di Francesco Maesa, P. Mori, and L. Ricci, "A blockchain based approach for the definition of auditable access control systems," *Comput. Secur.*, vol. 84, pp. 93–119, Jul. 2019.

[46] T. Liu, X. Chen, J. Li, S. Wu, W. Sun, and Y. Lu, "Research on progress of blockchain access control," in *Proc. IEEE 6th Int. Conf. Data Sci. Cyberspace (DSC)*, Dec. 2021, pp. 516–522.

[47] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, pp. 612–613, Nov. 1979, doi: 10.1145/359168.359176.

[48] T. P. Pedersen, "A threshold cryptosystem without a trusted party," in *Advances in Cryptology—EUROCRYPT'91*, D. W. Davies, Ed. Berlin, Germany: Springer, 1991, pp. 522–526.

[49] E. Hesamifard, H. Takabi, and M. Ghasemi, "CryptoDL: Deep neural networks over encrypted data," 2017, *arXiv:1711.05189*.

[50] R. Gilad-Bachrach, N. Dowlin, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing, "CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy," in *Proc. 33rd Int. Conf. Mach. Learn.*, New York, NY, USA, M. F. Balcan and K. Q. Weinberger, Eds. vol. 48, Jun. 2016, pp. 201–210. [Online]. Available: https://proceedings.mlr.press/v48/gilad-bachrach16.html

[51] S. Obla, X. Gong, A. Aloufi, P. Hu, and D. Takabi, "Effective activation functions for homomorphic evaluation of deep neural networks," *IEEE Access*, vol. 8, pp. 153098–153112, 2020.

[52] J. Konečný, B. McMahan, and D. Ramage, "Federated optimization: Distributed optimization beyond the datacenter," 2015, *arXiv:1511.03575*.

[53] V. Drungilas, E. Vaičiukynas, M. Jurgelaitis, R. Butkienė, and L. Čeponienė, "Towards blockchain-based federated machine learning: Smart contract for model inference," *Appl. Sci.*, vol. 11, no. 3, p. 1010, Jan. 2021.

[54] U. Majeed, L. U. Khan, A. Yousafzai, Z. Han, B. J. Park, and C. S. Hong, "ST-BFL: A structured transparency empowered cross-silo federated learning on the blockchain framework," *IEEE Access*, vol. 9, pp. 155634–155650, 2021.

[55] Intel. *Intel Software Guard Extensions*. Accessed: Jun. 2015. [Online]. Available: https://software.intel.com/sites/default/files/332680-001.pdf

[56] S. Fei, Z. Yan, W. Ding, and H. Xie, "Security vulnerabilities of SGX and countermeasures: A survey," *ACM Comput. Surveys*, vol. 54, no. 6, pp. 1–36, Jul. 2021.

[57] J. Götzfried, M. Eckert, S. Schinzel, and T. Müller, "Cache attacks on Intel SGX," in *Proc. 10th Eur. Workshop Syst. Secur.* New York, NY, USA: Association for Computing Machinery, 2017, pp. 1–3.

[58] K. Murdock, D. Oswald, F. D. Garcia, J. Van Bulck, D. Gruss, and F. Piessens, "Plundervolt: Software-based fault injection attacks against Intel SGX," in *Proc. IEEE Symp. Secur. Privacy (SP)*, Jun. 2020, pp. 1466–1482.

[59] S. van Schaik, A. Kwong, D. Genkin, and Y. Yarom, "SGAxe: How SGX fails in practice," Univ. Michigan, Ann Arbor, MI, USA, Tech. Rep., 2020. [Online]. Available: https://sgaxe.com/files/SGAxe.pdf

[60] Y. Uchida, Y. Nagai, S. Sakazawa, and S. Satoh, "Embedding watermarks into deep neural networks," in *Proc. ACM Int. Conf. Multimedia Retr.* New York, NY, USA: Association for Computing Machinery, 2017, pp. 269–277, doi: 10.1145/3078971.3078974.

[61] J. Zhang, Z. Gu, J. Jang, H. Wu, M. P. Stoecklin, H. Huang, and I. Molloy, "Protecting intellectual property of deep neural networks with watermarking," in *Proc. Asia Conf. Comput. Commun. Secur.* New York, NY, USA: Association for Computing Machinery, 2018, pp. 159–172, doi: 10.1145/3196494.3196550.

[62] T. Liu, X. Xie, and Y. Zhang, "ZkCNN: Zero knowledge proofs for convolutional neural network predictions and accuracy," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* New York, NY, USA: Association for Computing Machinery, Nov. 2021, pp. 2968–2985, doi: 10.1145/3460120.3485379.

[63] H. Al-Breiki, M. H. U. Rehman, K. Salah, and D. Svetinovic, "Trustworthy blockchain oracles: Review, comparison, and open research challenges," *IEEE Access*, vol. 8, pp. 85675–85685, 2020.

[64] R. Feng, M. Liu, and N. Zhang, "A unified theory of decentralized insurance," *SSRN Electron. J.*, pp. 24–34, Jan. 2022.

[65] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, and A. Ray, "Training language models to follow instructions with human feedback," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 27730–27744.

[66] (2023). *The Collective Intelligence Project*. [Online]. Available: https://cip.org/whitepaper

[67] W. Zhang, Y. Yuan, Y. Hu, S. Huang, S. Cao, A. Chopra, and S. Huang, "A privacy-preserving voting protocol on blockchain," in *Proc. IEEE 11th Int. Conf. Cloud Comput. (CLOUD)*, Jul. 2018, pp. 401–408.

[68] S. Park, M. Specter, N. Narula, and R. L. Rivest, "Going from bad to worse: From internet voting to blockchain voting," *J. Cybersecurity*, vol. 7, no. 1, Feb. 2021, Art. no. tyaa025, doi: 10.1093/cybsec/tyaa025.

[69] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, and A. Askell, "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1877–1901.

[70] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," 2021, *arXiv:2112.10752*.

[71] J. Lin, R. Men, A. Yang, C. Zhou, Y. Zhang, P. Wang, J. Zhou, J. Tang, and H. Yang, "M6: Multi-modality-to-multi-modality multitask mega-transformer for unified pretraining," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, 2021, pp. 3251–3261.

[72] O. Sharir, B. Peleg, and Y. Shoham, "The cost of training NLP models: A concise overview," 2020, *arXiv:2004.08900*.

[73] S. M. Werner, D. Perez, L. Gudgeon, A. Klages-Mundt, D. Harz, and W. J. Knottenbelt, "SoK: Decentralized finance (DeFi)," 2021, *arXiv:2101.08778*.

[74] ChainAnalysis. (Jun. 27, 2022). *Dissecting the DAO: Web3 Ownership is Surprisingly Concentrated*. [Online]. Available: https://blog.chainalysis.com/reports/web3-daos-2022/

[75] X. Sun and C. Stasinakis, "Decentralization illusion in DeFi: Evidence from MakerDAO," AI21 Studio, Tech. Rep., 2021.

[76] C. Kim. (Jun. 8, 2019). *How Blockchain Voting is Supposed to Work (but in Practice Rarely Does)*. [Online]. Available: https://www.coindesk.com/markets/2019/06/08/how-blockchain-voting-is-supposed-to-work-but-in-practice-rarelydoes/

[77] I. M. P. Daian, T. Kell and A. Juels. (Jul. 2, 2018). *On-Chain Vote Buying and the Rise of Dark DAOs* https://hackingdistributed.com/2018/07/02/on-chain-vote-buying/

[78] R. Fritsch, M. Müller, and R. Wattenhofer, "Analyzing voting power in decentralized governance: Who controls DAOs?" 2022, *arXiv:2204.01176*.

[79] H. Mei, G. Li, J. Wu, and L. Zheng, "Privacy inference-empowered stealthy backdoor attack on federated learning under non-IID scenarios," 2023, *arXiv:2306.08011*.

[80] C. Dwork, "Differential privacy: A survey of results," in *Theory and Applications of Models of Computation*. Xi'an, China: Springer, Apr. 2008, pp. 1–19.

[81] G. Li, J. Wu, S. Li, W. Yang, and C. Li, "Multitentacle federated learning over software-defined industrial Internet of Things against adaptive poisoning attacks," *IEEE Trans. Ind. Informat.*, vol. 19, no. 2, pp. 1260–1269, Feb. 2023.

[82] W. Ou, J. Zeng, Z. Guo, W. Yan, D. Liu, and S. Fuentes, "A homomorphic-encryption-based vertical federated learning scheme for rick management," *Comput. Sci. Inf. Syst.*, vol. 17, no. 3, pp. 819–834, 2020.

[83] F. Mi, Z. Wang, C. Zhao, J. Guo, F. Ahmed, and L. Khan, "VSCL: Automating vulnerability detection in smart contracts with deep learning," in *Proc. IEEE Int. Conf. Blockchain Cryptocurrency (ICBC)*, May 2021, pp. 1–9.

[84] Z. Liu, P. Qian, X. Wang, L. Zhu, Q. He, and S. Ji, "Smart contract vulnerability detection: From pure neural network to interpretable graph feature and expert pattern fusion," in *Proc. 13th Int. Joint Conf. Artif. Intell.*, Z.-H. Zhou, Ed. Aug. 2021, pp. 2751–2759, doi: 10.24963/ijcai.2021/379.

[85] J. Ye, M. Ma, Y. Lin, L. Ma, Y. Xue, and J. Zhao, "Vulpedia: Detecting vulnerable Ethereum smart contracts via abstracted vulnerability signatures," *J. Syst. Softw.*, vol. 192, Oct. 2022, Art. no. 111410. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0164121222001236

[86] P. Tolmach, Y. Li, S.-W. Lin, Y. Liu, and Z. Li, "A survey of smart contract formal specification and verification," *ACM Comput. Surveys*, vol. 54, no. 7, pp. 1–38, Jul. 2021, doi: 10.1145/3464421.

[87] T. Vaiyapuri, K. Shankar, S. Rajendran, S. Kumar, S. Acharya, and H. Kim, "Blockchain assisted data edge verification with consensus algorithm for machine learning assisted IoT," *IEEE Access*, vol. 11, pp. 55370–55379, 2023.

[88] H. Kim, S.-H. Kim, J. Y. Hwang, and C. Seo, "Efficient privacy-preserving machine learning for blockchain network," *IEEE Access*, vol. 7, pp. 136481–136495, 2019.

[89] GDPR. *Complete Guide to GDPR Compliance*. Accessed: Jan. 10, 2021. [Online]. Available: https://gdpr.eu/

[90] State of California Department of Justice. *California Consumer Privacy Act (CCPA)*. Accessed: Jan. 21, 2020. [Online]. Available: https://oag.ca.gov/privacy/ccpa
[91] National Conference of State Legislatures. (Aug. 26, 2022). *Legislation Related to Artificial Intelligence*. [Online]. Available: https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx
[92] A. Pollacco. (2020). *The Interaction Between Blockchain Evidence and Courts*. [Online]. Available: https://blog.bcas.io/blockchain_court_evidence

DANA ALSAGHEER received the B.S. degree in computer science from the University of Houston, in 2019, where she is currently pursuing the Ph.D. degree. Her research interests include machine learning and blockchain.

LEI XU is currently an Assistant Professor with Kent State University. His research interests include blockchain, cloud security, and applied cryptography.

WEIDONG SHI (Senior Member, IEEE) is currently an Associate Professor with the University of Houston. His research interests include high-performance computer architecture for large-scale multimedia applications, applied security, real-time computer graphics, blockchain, and cloud computing.