On Decentralized Governance of Machine Learning and AI

Dana Alsagheer*, Nour Diallo*, Rabimba Karanjai*, Lei Xu†, Larry Shi*
*University Of Houston, TX, USA
{dralsagh,rkaranjai, ndiallo, mkaleem, wshi3}@uh.edu
†Kent State University, OH, USA ,
xuleimath@gmail.com

***Abstract*—In recent years, machine learning (ML)and artificial intelligence (AI) have enabled various applications, such as data analytics and autonomous systems, including blockchain-based technology. ML and AI are now pervasive new systems and models are being deployed in every domain imaginable, leading to widespread software-based inference and decision-making. Researchers recognize the effectiveness of approaches to a well-defined ML and AI governance framework based on the principle of decentralization and comprehensively define its scope of research and practice. In this paper, we study ML and AI value chain management, decentralized identity for the ML community, and ownership and rights management of ML assets (data, model, code). Furthermore, community-based decision-making for the ML process, decentralized ML finance, and risk management.**

***Index Terms*—machine learning, decentralization, governance, blockchain, DAO**

I. Introduction

In 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-based 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 applied to and validated by the success of applications such as DAOs, DeFi governance [5], and governance of blockchain protocols.

In this work, we expand the concept of ML governance under the lens of decentralization. We proposed a new framework of decentralized ML governance that encompasses 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 many new research topic areas like community-owned and community-managed ML process, 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 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 component;*

(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. Decentralized ML Governance

In this work, 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], [6], [7]. 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. The Scope of Decentralized ML

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 be applied to 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 Cocreation

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 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 because 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, model 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 pipeline view of the ML process.

The transformative power of casting ML governance as a value chain process is that it makes integrating ML

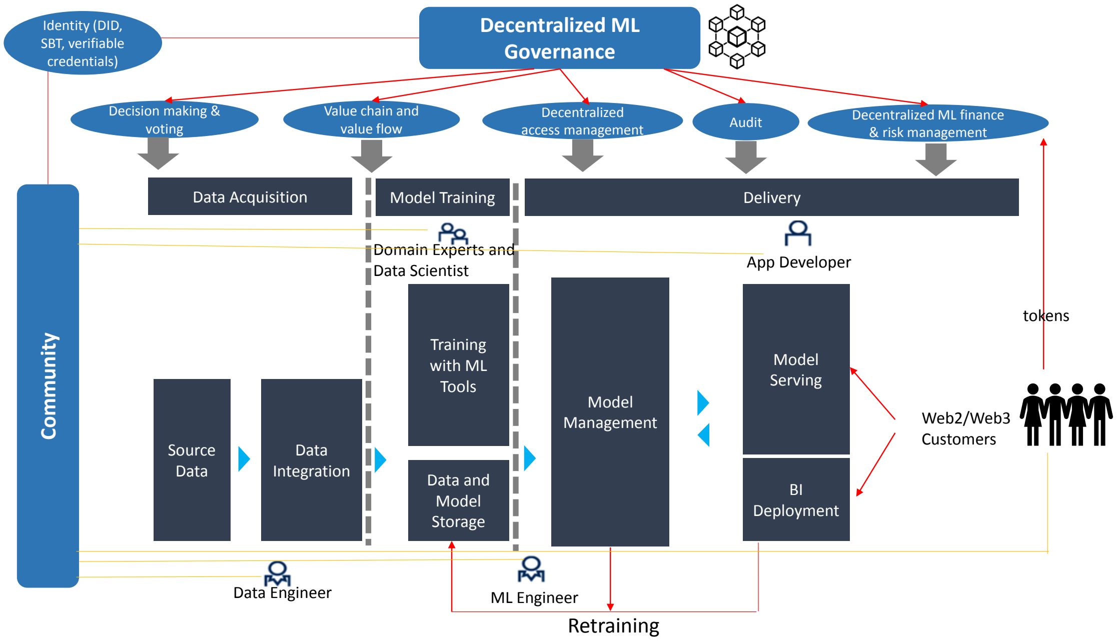

Figure 1: Decentralized ML Governance. The components (gray boxes) are defined by MLOps. Decentralized ML governance focuses on the blue components and the seamless integration of these components with the gray boxes.

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 [8], [9]. 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 [10], [11], 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 transparency of the value creation process and the open coordination among the stakeholders make the approach attractive. There are many options for distributing ML values among the stakeholders, like revenue sharing 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 trust among the stakeholders and the participants. Once joining the ML projects, the participants are incentivized to engage and collaborate closely in the ML value-creation process.

D. *Decentralized Identity*

For ML governance, digital identity is a fundamental component. 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 of 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

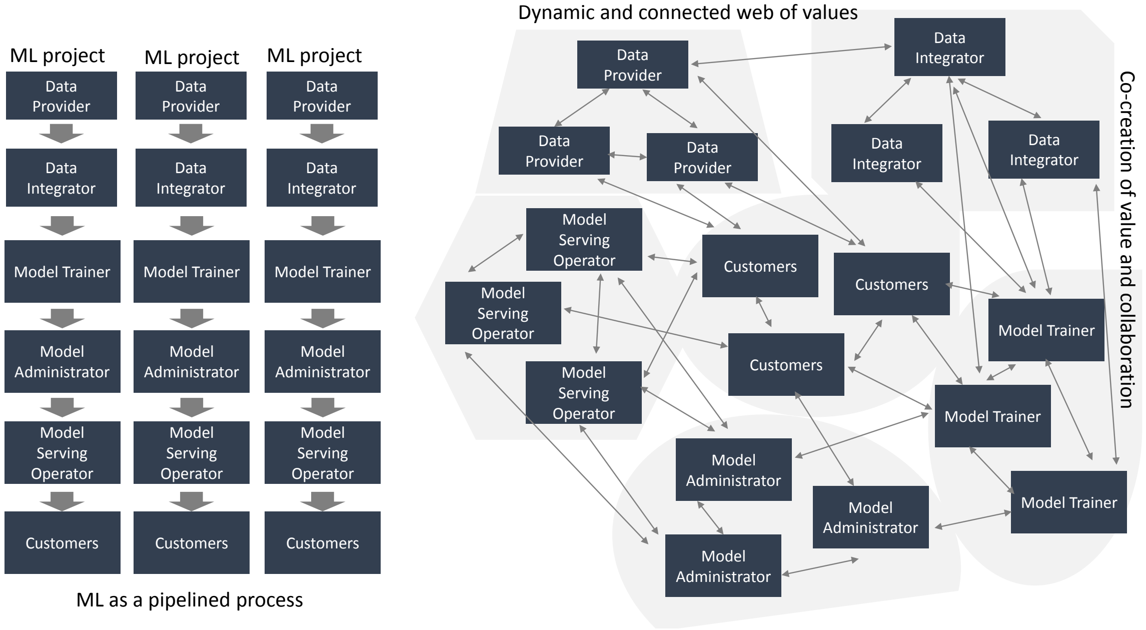

Figure 2: Pipeline view of ML process vs. view of ML governance as a value web with co-creation by the stakeholder community.

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 [12], OpenID [13]). To support interoperability and avoid fragmentation, federated identity management [14] 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 has emerged in recent years, called self-sovereign identities (SSI) [15]. 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, in SSI, the 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) [16]. According to W3C, a DID is a digital identifier that does not need 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 [17] 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 [18]. For instance, stakeholders can make claims like 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) [19] has been 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, a soul-bound token can be certified by other ML specialists or users who interact with it in a ML community. 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.

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, right 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 rights to the ML artifacts to other ML participants. 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 a 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 the usage of 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?

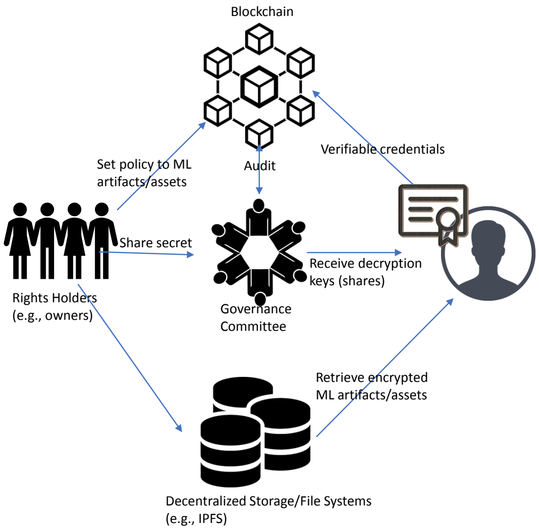
Blockchain
Set policy to ML artifacts/assets
Audit
Verifiable credentials
Share secret
Rights Holders (e.g., owners)
Governance Committee
Receive decryption keys (shares)
Retrieve encrypted ML artifacts/assets
Decentralized Storage/File Systems (e.g., IPFS)

Figure 3: Decentralized access management of ML assets/artifacts.

Prior work exists that attempts to leverage blockchains and distributed ledger technology for digital rights management [20], [21], [22]. 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 the 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., [23], [24]). 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 for eventual development of 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 [25], [26]. 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 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 they are in use (e.g., training and inference). There are several technology frontiers under active research and development to provide such solutions. Homomorphic ML is one area where ML tasks can be performed over encrypted data [27], [28]. Despite of 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 [29] because of the computation cost. To protect the confidentiality of the training data, federated ML is a promising direction to explore [30]. Researchers have started integrating federated ML with smart contracts and blockchains [31], [32].

Another direction is to leverage hardware with specific security features, such as Trusted Execution Environments (TEEs), for ML (e.g., [33]). Compared with other options like homomorphic ML, TEE-based ML can deliver better performance. However, TEE-based approaches face their own challenges, such as lack of vendor-agnostic standards in TEE implementation; low performance compared with non-TEE based MLvulnerabilities to attacks like side-channel exploits and other exposed attack surfaces (e.g., [34], [35], [36], [37]).

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 verification 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 [38], [39], 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 [40], or ML-oriented verifiable computations where computation and transformation applied to ML models can be publicly verified.

F. Opportunities of Decentralized Risk Management

ML process, by nature, involves risks. The risks can be security 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 have implications for 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 towards 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 [41] 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 approaches to managing risks: 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 staking for managing trust and risks. However, staking has its downsides, for instance, high cost to the service providers and efficiency issues due to the lock of financial assets during staking. Decentralized insurance [42] 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.

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.

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 both 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 overall 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 [43]), defense against attacks and manipulation of the decision-making/voting process (e.g., [44], mitigation of governance risks, fairness in the governance process).

H. Decentralized ML Finance

ML finance is a necessary part of ML governance. Trained ML models are at the center of ML systems because of their potential to support the diverse and large number of impactful applications (e.g., GPT-3 [45], NLLB-200 [46], stable diffusion model [47], M6 model by Damo Academy [48]). 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 the 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 [49], 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 in the future.

To solve the ML financing challenge, decentralized ML governance can benefit from the rich space of decentralized finance [5]. 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?

III. Challenges

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 size 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., [50], [51], [52]). 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 [53]. 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 very few participants who have more resources than the others in the community [50], [54]. 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 is implemented as smart contract code and/or off-chain software that operates in tandem with on-chain code. The off-chain and on-chain code may have security vulnerabilities that bad actors could exploit (e.g., [55]). Despite the recent advance in smart contract audit, automated vulnerability detection, and formal verification of smart contract properties (e.g., [56], [57], [58], [59]), it is impossible to guarantee that governance software is completely vulnerability free. Successful exploits of DAOs could lead to serious damage to the interests of the stakeholder community both in terms of reputation, trust in the system, and finance.

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 standard could be developed to provide a solution for a sub-problem within each area, for instance, the standard for decentralized identity or decentralized management of digital rights. A challenge is integrating the results and research outcomes in each area into a complete solution for ML governance with interoperability. 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 ML governance needs to solve many security/privacy challenges related to identity, ownership and rights management, access control, voting, audit, and ML with privacy and confidentiality assurance. At the same time, decentralized governance with privacy guarantees should not undermine the accountability and trustworthiness of decentralized governance. For instance, privacy protection allows stakeholders to participate in governance without revealing their real identities, and this can become a double-edged sword to the long-term well-being of decentralized governance because bad actors can take advantage of the strong privacy protection to engage in unethical or even unlawful actions. Without proper management and design of the governance mechanism based on privacy-preserving technologies, privacy could result in a lack of accountability and hinder the wide acceptance of decentralized 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 [60] and CCPA (California Consumer Privacy Act) [61] 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 at ML and AI [62]. For centralized and decentralized ML governance, compliance and legal audit would be significant challenges from both technical and legal aspects. For instance, how a decentralized organization manages compliance requirements, responds to the requests from regulatory agencies, and operates to meet all the consumer demands in a timely manner. 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.

IV. Related Work

Decentralized governance. With the advance of blockchain technology and virtual organizations like DAOs [3], [6], researchers and practitioners have explored the applications of blockchains and DAOs for decentralized governance. For example, many blockchains implement their governance mechanisms [4]. Most popular DeFi projects have adopted decentralized governance for managing stakeholder communities. These existing decentralized governance designs are not specifically developed for the ML industry. In this work, we focus on exclusively defining the scope of decentralized governance for ML.

ML governance. The concept of ML governance has been studied within the ML community (e.g., [2]). However, the prior efforts in this area primarily focus on security, privacy, and compliance-related issues. The defined scope of governance is significantly different from what this paper describes. The introduction of concepts such as ML value chain management, community-owned, and managed ML assets, economic and financial risk management for ML, and decentralized ML finance, distinguishes this work from the prior endeavor in defining ML governance. In other words, this work drastically broadens the scope of ML governance in general and delineates decentralized ML governance in particular.

MLOps. MLOps combines ML process with DevOps [63]. As discussed earlier, the scope of MLOps is much narrower than what is presented here [64]. MLOps mainly target at enterprise customers and rely on a centralized orchestration model. The framework in this paper is based on the principle of decentralization, which is applied comprehensively to various parts of ML governance (e.g., identity, ownership, rights management, decision-making, value chain management, finance, and risk management).

V. Conclusions

Machine learning in computer systems introduces many benefits but also raises risks to society which indicates 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 the 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, and community management. In this paper, we study in depth the details of the decentralized ML governance framework and provide a comprehensive view of its components, research opportunities, challenges, and open problems.

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, pp. 706–710, 2020.

[2] V. Chandrasekaran, H. Jia, A. Thudi, A. Travers, M. Yaghini, and N. Papernot, "Sok: Machine learning governance," *CoRR*, vol. abs/2109.10870, 2021. [Online]. Available: https://arxiv.org/abs/2109.10870

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

[4] R. van Pelt, S. Jansen, D. Baars, and S. Overbeek, "Defining blockchain governance: A framework for analysis and comparison," *Information Systems Management*, vol. 38, no. 1, pp. 21–41, 2021. [Online]. Available: https://doi.org/10.1080/10580530.2020.1720046

[5] S. M. Werner, D. Perez, L. Gudgeon, A. Klages-Mundt, D. Harz, and W. J. Knottenbelt, "Sok: Decentralized finance (defi)," 2021. [Online]. Available: https://arxiv.org/abs/2101.08778

[6] C. Hackly, "What are daos and why you should pay attention," https://www.forbes.com/sites/cathyhackl/2021/06/01/what-are-daos-and-why-you-should-pay-attention/?sh=1f879a467305, 1 June 2021.

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

[8] C. Grönroos and A. Ravald, "Service as business logic: implications for value creation and marketing," *Journal of Service Management*, vol. 22, pp. 5–22, 2011.

[9] S. L. Vargo and R. F. Lusch, "Institutions and axioms: an extension and update of service-dominant logic," *Journal of the Academy of Marketing Science*, vol. 44, no. 1, pp. 5–23, January 2016. [Online]. Available: https://ideas.repec.org/a/spr/joamsc/v44y2016i1d10.1007_s11747-015-0456-3.html

[10] M. Mačiulienė and A. Skaržauskienė, "Conceptualizing blockchain-based value co-creation: A service science perspective," *Systems Research and Behavioral Science*, vol. 38, no. 3, pp. 330–341, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sres.2786

[11] E. Seulliet, "Open innovation, co-creation: Why blockchain is a small revolution," https://medium.com/@ericseulliet/open-innovation-co-creation-whyblockchain-is-asmall-revolution-73e7d0b480d5, 2016.

[12] D. Hardt, "The oauth 2.0 authorization framework," Internet Requests for Comments, RFC Editor, RFC 6749, October 2012, http://www.rfc-editor.org/rfc/rfc6749.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc6749.txt

[13] N. Sakimura, J. Bradley, and M. Jones, "Openid connect dynamic client registration 1.0 incorporating errata set 1. openid foundation," http://openid.net/specs/openid-connect-registration-1_0.html, 2014.

[14] D. W. Chadwick, *Federated Identity Management.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 96–120. [Online]. Available: https://doi.org/10.1007/978-3-642-03829-7_3

[15] R. Soltani, U. T. Nguyen, A. An, and C. Galdi, "A survey of self-sovereign identity ecosystem," *Sec. and Commun. Netw.*, vol. 2021, jan 2021. [Online]. Available: https://doi.org/10.1155/2021/8873429

[16] W3C, "Decentralized identifiers (dids) v1.0. core architecture, data model, and representations," https://www.w3.org/TR/2020/WD-did-core-20201108/, 08 November 2020.

[17] M. Sporny, G. Noble, D. Longley, D. C. Burnett, B. Zundel, and K. D. Hartog, "Verifiable credentials data model 1.0," https://www.w3.org/TR/verifiable-claims-data-model/, 2019.

[18] U. Feige, A. Fiat, and A. Shamir, "Zero-knowledge proofs of identity," *Journal of cryptology*, vol. 1, no. 2, pp. 77–94, 1988.

[19] M. Zoltu, "Eip-5114: Soulbound badge," https://ethereum-magicians.org/t/eip-5114-soulbound-token/9417, 05 May 2022.

[20] Z. Zhang and L. Zhao, "A design of digital rights management mechanism based on blockchain technology," in *International Conference on Blockchain*, 2018.

[21] M. Hasan, M. Tariq, Z. Chen, A. Dwivedi, M. Kamal, A. Garba Ph.D., and G. Srivastava, "A digital rights management system based on a scalable blockchain," *Peer-to-Peer Networking and Applications*, 09 2021.

[22] M. Holland, J. Stjepandić, and C. Nigischer, "Intellectual property protection of 3d print supply chain with blockchain technology," in *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2018, pp. 1–8.

[23] "A blockchain based approach for the definition of auditable access control systems," *Computers & Security*, vol. 84, pp. 93–119, 2019.

[24] T. Liu, X. Chen, J. Li, S. Wu, W. Sun, and Y. Lu, "Research on progress of blockchain access control," in *2021 IEEE Sixth International Conference on Data Science in Cyberspace (DSC)*, 2021, pp. 516–522.

[25] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, p. 612–613, nov 1979. [Online]. Available: https://doi.org/10.1145/359168.359176

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

[27] E. Hesamifard, H. Takabi, and M. Ghasemi, "Cryptodl: Deep neural networks over encrypted data," 2017. [Online]. Available: https://arxiv.org/abs/1711.05189

[28] 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 *Proceedings of The 33rd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. F. Balcan and K. Q. Weinberger, Eds., vol. 48. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 201–210. [Online]. Available: https://proceedings.mlr.press/v48/gilad-bachrach16.html

[29] S. Obla, "Effective activation functions for homomorphic evaluation of deep neural networks," 2020.

[30] J. Konečný, B. McMahan, and D. Ramage, "Federated optimization: Distributed optimization beyond the datacenter," *CoRR*, vol. abs/1511.03575, 2015. [Online]. Available: http://arxiv.org/abs/1511.03575

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

[32] 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. 155 634–155 650, 2021.

[33] Intel, "Intel software guard extensions," https://software.intel.com/sites/default/files/332680-001.pdf.

[34] S. Fei, Z. Yan, W. Ding, and H. Xie, "Security vulnerabilities of SGX and countermeasures: A survey," *ACM Computing Survey*, vol. 54, no. 6, Jul. 2021.

[35] J. Götzfried, M. Eckert, S. Schinzel, and T. Müller, "Cache attacks on intel sgx," in *Proceedings of the 10th European Workshop on Systems Security*, ser. EuroSec'17. New York, NY, USA: Association for Computing Machinery, 2017.

[36] 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 *2020 IEEE Symposium on Security and Privacy (S&P)*, 2020, pp. 1466–1482.

[37] S. van Schaik, A. Kwong, D. Genkin, and Y. Yarom, "Sgaxe: How sgx fails in practice," 2020.

[38] Y. Uchida, Y. Nagai, S. Sakazawa, and S. Satoh, "Embedding watermarks into deep neural networks," in *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, ser. ICMR '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 269–277. [Online]. Available: https://doi.org/10.1145/3078971.3078974

[39] 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 *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, ser. ASIACCS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 159–172. [Online]. Available: https://doi.org/10.1145/3196494.3196550

[40] T. Liu, X. Xie, and Y. Zhang, "Zkcnn: Zero knowledge proofs for convolutional neural network predictions and accuracy," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 2968–2985. [Online]. Available: https://doi.org/10.1145/3460120.3485379

[41] 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. 85 675–85 685, 2020.

[42] R. Feng, M. Liu, and N. Zhang, "A unified theory of decentralized insurance," *SSRN Electronic Journal*, 01 2022.

[43] W. Zhang, Y. Yuan, Y. Hu, S. Huang, S. Cao, A. Chopra, and S. Huang, "A privacy-preserving voting protocol on blockchain," in *2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*, 2018, pp. 401–408.

[44] S. Park, M. Specter, N. Narula, and R. L. Rivest, "Going from bad to worse: from Internet voting to blockchain voting," *Journal of Cybersecurity*, vol. 7, no. 1, 02 2021, tyaa025. [Online]. Available: https://doi.org/10.1093/cybsec/tyaa025

[45] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," 2020. [Online]. Available: https://arxiv.org/abs/2005.14165

[46] NLLB Team, M. R. Costa-jussà, J. Cross, O. Çelebi, M. Elbayad, K. Heafield, K. Heffernan, E. Kalbassi, J. Lam, D. Licht, J. Maillard, A. Sun, S. Wang, G. Wenzek, A. Youngblood, B. Akula, L. Barrault, G. M. Gonzalez, P. Hansanti, J. Hoffman, S. Jarrett, K. R. Sadagopan, D. Rowe, S. Spruit, C. Tran, P. Andrews, N. F. Ayan, S. Bhosale, S. Edunov, A. Fan, C. Gao, V. Goswami, F. Guzmán, P. Koehn, A. Mourachko, C. Ropers, S. Saleem, H. Schwenk, and J. Wang, "No language left behind: Scaling human-centered machine translation," 2022. [Online]. Available: https://arxiv.org/abs/2207.04672

[47] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," *CoRR*, vol. abs/2112.10752, 2021. [Online]. Available: https://arxiv.org/abs/2112.10752

[48] 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 *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 3251–3261.

[49] O. Sharir, B. Peleg, and Y. Shoham, "The cost of training nlp models: A concise overview." *CoRR*, vol. abs/2004.08900, 2020. [Online]. Available: http://dblp.uni-trier.de/db/journals/corr/corr2004.html#abs-2004-08900

[50] ChainAnalysis, “Dissecting the dao: Web3 ownership is surprisingly concentrated,” https://blog.chainalysis.com/reports/web3-daos-2022/, June 27, 2022.

[51] X. Sun, C. Stasinakis, and G. Sermpinis, “Decentralization illusion in defi: Evidence from makerdao,” 2022. [Online]. Available: https://arxiv.org/abs/2203.16612

[52] C. Kim, “How blockchain voting is supposed to work (but in practice rarely does),” https://www.coindesk.com/markets/2019/06/08/how-blockchain-voting-is-supposed-to-work-but-in-practice-rarelydoes/, June 8 2019.

[53] I. M. Philip Daian, Tyler Kell and A. Juels, “On-chain vote buying and the rise of dark daos,” https://hackingdistributed.com/2018/07/02/on-chain-vote-buying/, July 02, 2018.

[54] R. Fritsch, M. Müller, and R. Wattenhofer, “Analyzing voting power in decentralized governance: Who controls daos?” 2022. [Online]. Available: https://arxiv.org/abs/2204.01176

[55] N. Atzei, M. Bartoletti, and T. Cimoli, “A survey of attacks on ethereum smart contracts sok,” in *Proceedings of the 6th International Conference on Principles of Security and Trust - Volume 10204*. Berlin, Heidelberg: Springer-Verlag, 2017, p. 164–186. [Online]. Available: https://doi.org/10.1007/978-3-662-54455-6_8

[56] F. Mi, Z. Wang, C. Zhao, J. Guo, F. Ahmed, and L. Khan, “Vscl: Automating vulnerability detection in smart contracts with deep learning,” in *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 2021, pp. 1–9.

[57] 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 *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, Z.-H. Zhou, Ed. International Joint Conferences on Artificial Intelligence Organization, 8 2021, pp. 2751–2759, main Track. [Online]. Available: https://doi.org/10.24963/ijcai.2021/379

[58] J. Ye, M. Ma, Y. Lin, L. Ma, Y. Xue, and J. Zhao, “Vulpedia: Detecting vulnerable ethereum smart contracts via abstracted vulnerability signatures,” *Journal of Systems and Software*, vol. 192, p. 111410, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0164121222001236

[59] P. Tolmach, Y. Li, S.-W. Lin, Y. Liu, and Z. Li, “A survey of smart contract formal specification and verification,” *ACM Comput. Surv.*, vol. 54, no. 7, jul 2021. [Online]. Available: https://doi.org/10.1145/3464421

[60] GDPR, “Complete guide to gdpr compliance,” https://gdpr.eu/, last accessed on 10/1/2021.

[61] S. of California Department of Justice, “California consumer privacy act (ccpa),” https://oag.ca.gov/privacy/ccpa, last accessed on 01/21/2020.

[62] N. C. of State Legislatures, “Legislation related to artificial intelligence,” https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx, 26, Aug 2022.

[63] D. Kreuzberger, N. Kühl, and S. Hirschl, “Machine learning operations (mlops): Overview, definition, and architecture,” 2022. [Online]. Available: https://arxiv.org/abs/2205.02302

[64] D. Kreuzberger, N. Kühl, and S. Hirschl, “Machine learning operations (mlops): Overview, definition, and architecture,” *arXiv preprint arXiv:2205.02302*, 2022.