Decentralized Machine Learning Governance

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

2023 IEEE International Conference on Blockchain and Cryptocurrency (ICBC) | 979-8-3503-1019-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICBC56567.2023.10175000

***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 for 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.**

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.

II. Decentralized ML Governance

A. 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 entities in the ML pipeline where each entity is specialized in providing services of one stag.

casting machine learning (ML) governance as a value chain process can facilitate its integration with blockchains and Web3. Value co-creation, which involves collaborative processes with stakeholders and end-users, is a natural fit with blockchain's properties like traceability, transparency, and capitalization of transactions. In decentralized value co-creation, autonomous ML value chain stakeholders can collaborate for a specific ML project or system and distribute revenue and income among themselves using a previously agreed upon on-chain distribution model. The use of smart contracts can automate the execution of the value chain process, improve trust among stakeholders, and incentivize participation and collaboration in the ML value-creation process.

B. Decentralized Identity

Digital identitis necessary to access ML artifacts, resources, and services, participate in ML governance, and make operational decisions.

Following some identity management :

- Identity Access Management(IAM) 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 [6], OpenID [7]). To support interoperability and avoid fragmentation, federated identity management [8] is developed to enable users to access the resources and services of multiple organizations using a single set of credentials.
- Self-sovereign identities (SSI) [9] are more decentralized and based on blockchains. It puts end-users entirely in control and allows different service providers to share identity verification attestations.

979-8-3503-1019-1/23/$31.00 ©2023 IEEE

- Decentralized Identifiers(DID) is the W3C initiative on standardizing DIDs [10]. 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 [11] to identify people, organizations, and things to achieve desired security and privacy-protection guarantees.
- Soul-bound tokens (SBT) [12] has been proposed to achieve the vision of a decentralized society. Soul-bound tokens are publicly visible and non-transferable tokens.

C. Opportunities of Decentralized Risk Management

The ML process involves security, privacy, and societal and financial risks. ML models can face attacks and have poor robustness, while their fairness can affect social justice. There is a need for service level agreements (SLA) for ML services and systems to hold service providers accountable for potential damages to end users. The current economic bias toward service providers may hinder society's adoption of ML and affect trust. ML can have applications such as DeFi, business processing, dApps, and 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 use three methods to manage risks: reputation-based, staking-based, and insurance-based. Staking is suitable for a fully decentralized and permissionless environment where service providers are anonymous. However, staking has downsides, such as high provider costs and efficiency issues due to a lack of financial assets during staking. Many dApps use staking for managing trust and risks. Decentralized insurance [13] and risk management are attractive because these approaches can lower the cost for the stakeholders.

D. 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 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 [14]), defense against attacks and manipulation of the decision-making/voting process (e.g., [15], mitigation of governance risks, fairness in the governance process.

E. 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 a diverse and large number of impactful applications (e.g., GPT-3 [16], NLLB-200 [17], stable diffusion model [18], M6 model by Damo Academy [19]). 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. 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.

III. Conclusions

Decentralized ML governance leverages technologies such as blockchains, distributed ledgers, and smart contracts to support broader ML governance beyond security and privacy, including value chain management, ML finance, and community management.

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] 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

[7] 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.

[8] 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

[9] 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

[10] 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.

[11] 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.

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

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

[14] 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.

[15] 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

[16] 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

[17] 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

[18] 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

[19] 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.