AI GOVERNANCE VIA WEB3 REPUTATION SYSTEM

Wulf Kaal*

ABSTRACT

In the rapidly evolving landscape of artificial intelligence (AI), governance frameworks are increasingly crucial. As AI technologies become more complex and integral to various sectors, the mechanisms to oversee and regulate these systems must evolve correspondingly. Traditional governance approaches often rely on static, predefined rules that may not adapt quickly enough to the pace of AI development or the nuanced challenges it presents. These conventional methods, largely reactive or fixed to ex-post solutions, are proving insufficient for the dynamic nature of AI technologies.

The proposed AI governance system integrates decentralized web3 community governance with federated communication platforms, forming a sophisticated framework for dynamic, anticipatory, and participatory oversight of AI development. Key components include a federated forum platform structured as a Weighted Directed Acyclic Graph (WDAG) and specialized smart contracts for managing tasks and validation. This setup facilitates real-time consensus-building and decision-making, supports a scalable, transparent communication network, and maintains an updated and responsive governance system through Validation Pools and Reputation tokens.

The WDAG system provides a dynamic, structured approach to AI governance, ensuring that models consistently align with evolving legal and ethical standards. Each AI model is represented as a node within the WDAG, with directed edges connecting it to relevant legal precedents and ethical guidelines. These connections are weighted to reflect the importance of each governance element, allowing real-time adjustments as new regulations or ethical considerations emerge, ensuring continuous compliance within diverse operational environments.

---

* Professor of Law, University of St. Thomas. The author is grateful for excellent support by research assistant Jack Palmer and research librarian Adam Bent.

INTRODUCTION

In the rapidly evolving landscape of artificial intelligence (AI), governance frameworks are increasingly pivotal. As AI technologies become more complex and integral to various sectors, the mechanisms to oversee and regulate these systems must evolve correspondingly. Traditional governance approaches often rely on static, predefined rules that may not adapt quickly enough to the pace of AI development or the nuanced challenges it presents. These conventional methods, largely reactive or fixed to ex-post solutions, are proving insufficient for the dynamic nature of AI technologies.

Web3, characterized by its decentralized architecture and reliance on blockchain technologies, introduces a fundamentally different approach to AI governance. This approach is not just a shift in technology but a paradigm shift in how governance can be implemented—namely, through decentralized autonomous governance systems that utilize dynamic feedback mechanisms.[1] Such systems are crucial to address the dual needs when governing evolving AI models ex-ante and managing existing solutions ex-post. No legacy systems exist at the time of publication that could provide such dynamic governance toolsets.

The traditional ex-post governance methods, where regulations are applied after AI systems are developed and deployed, or were pretrained LLMs models were trained on existing proprietary datasets, often fall short in preemptively addressing risks and biases. This approach can lead to gaps in oversight during critical early stages of AI development, where foundational attributes of AI systems are established. In contrast, an ex-ante governance approach, advocated within Web3 frameworks as presented in this paper, involves setting community coordinated regulatory measures and oversight mechanisms during the development phase of AI technologies. This proactive stance allows for real-time adjustments and refinements based on ongoing feedback from AI operations and interactions within the ecosystem.

---

[1] FESTSCHRIFT ZU EHREN VON CHRISTIAN KIRCHNER: RECHT IM ÖKONOMISCHEN KONTEXT XVI, 1387 (Wulf A. Kaal, Andreas Schwartze & Matthias Schmidt eds., 2014) [hereinafter FESTSCHRIFT].

Web3 systems, with their inherent capabilities for real-time data processing, decentralized decision-making, and transparent operationalizing, provide a robust infrastructure for implementing AI governance systems ex-ante. These systems facilitate a dynamic feedback loop, where AI behaviors and outcomes are continually monitored and influenced by decentralized consensus rather than solely dictated by centralized authorities or delayed regulatory responses. Technology has historically outpaced regulation.[2] The exponential trends in AI development will continue to exacerbate the mismatch between regulation and AI development. The author has advocated ex-ante dynamic regulatory methods for over a decade.[3] The proposed model in this paper ensures that AI governance is more adaptive, responsive, and aligned with ethical standards and societal needs ex-ante and during AI model development, not ex-post as demanded by traditional regulatory methods.

Moreover, the use of dynamic feedback mechanisms enables a more nuanced and effective management of AI technologies. By integrating feedback directly into the governance processes, stakeholders can iteratively improve and adjust AI models in response to new information, operational experiences, and changing environments. This ongoing process helps mitigate risks and biases more effectively than static, ex-post regulatory frameworks.

The shift towards Web3 and the utilization of Web3 tools with dynamic feedback effects represents a necessary advancement in AI governance frameworks. This new model not only addresses the inadequacies of previous systems but also enhances the capability to govern AI technologies in a manner that is as advanced and dynamic as the technologies themselves. This paper aims to explore these advancements in detail, illustrating how Web3 can transform AI governance from a reactive to a proactive discipline that is better equipped to handle the complexities of modern AI systems.

The conceptual framework for AI governance within Web3 revolves around creating a system where AI entities can operate autonomously yet responsibly within a decentralized digital ecosystem. In Web3, AI entities can perform transactions without human intervention. This is made possible through the use of cryptocurrencies and digital tokens. AI systems can have their wallets to send and receive tokens as payment for services, such as data

---

[2] Mark Fenwick, Wulf A. Kaal & Erik P.M. Vermeulen, *Regulation Tomorrow: What Happens When Technology Is Faster Than the Law?*, 6 AM. U. BUS. L. REV. 561, 564 (2017), https://ssrn.com/abstract=3204119.

[3] Mark Fenwick, Wulf A. Kaal, Toshiyuki Kono & Erik P.M. Vermeulen*, Dynamic Regulation for Innovation*, 6 PERSP. L., BUS. & INNOVATION 133, 135 (2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2831040.

processing or cognitive tasks. This capability is critical for fostering a self-sustaining ecosystem of AI services. Yet, the web3 technology enables human inputs to co-govern with AI in real-time. With the massive amounts of data that AI systems generate and process, ensuring the integrity and authenticity of this data is crucial.

Data attestation mechanisms in Web3 offer a cryptographic means to verify data integrity, which is crucial for AI systems reliant on data inputs for decision-making and pattern recognition. This capability not only enables AI tracking to monitor uploaded content and to address potential intellectual property concerns but also provides a governance framework to mitigate liability issues stemming from AI misuse. By ensuring accountability at the individual level, rather than burdening the entire community, this proposed model for AI governance within Web3 offers a promising avenue to address ethical dilemmas associated with AI applications.

Web3-based self-sovereign identity (SSI) allows individuals—and potentially AI systems—to have control over their digital identities without relying on a central authority. In the context of AI governance, SSI ensures that AI entities have verifiable credentials and can engage in interactions with confidence in the identity and reputation of other participants. Web3 can offer a set of standards and protocols that define how AI systems interact, share data, and settle transactions. These protocols act as the "rules of the road," setting the groundwork for interoperability and collaborative efforts among diverse AI systems.

In a decentralized autonomous organization (DAO) governed by Web3 principles, AI systems can actively participate in the governance process, contributing to decision-making and protocol changes that impact the network. This approach ensures a fair distribution of power, preventing any single entity from exerting undue influence and fostering an environment where decisions benefit all participants. Despite concerns about AI involvement in governance, the potential benefits are significant, especially when considering specialized AI systems that offer expertise or ensure the integrity of organizational principles through formal logic representation. By leveraging AI capabilities, DAOs can enhance the purity of decision-making processes, minimizing biases and agenda-driven actions, thus advancing the overall effectiveness and transparency of governance within Web3 ecosystems.

By integrating these features, Web3 aims to augment human intelligence rather than replace it. The goal is to create an environment where AI can process vast amounts of data, learn, and interact in ways that amplify our cognitive capabilities, leading to better decision-making and innovation. AI governance in Web3 is about creating a framework where AI systems are both empowered to act autonomously and accountable to the larger network. This involves enabling secure transactions, ensuring data

integrity, managing identities, and coordinating actions through shared protocols. By doing so, Web3 can foster a governance model that is adaptable, transparent, and responsive to the fast-paced changes characteristic of modern AI technologies.

1. ORIGIN OF AI

AI represents a fusion of computational techniques engineered to execute tasks traditionally requiring human intelligence, such as learning, reasoning, problem-solving, perception, and language understanding. The foundational architecture of AI systems incorporates a variety of methodologies that enable these machines to analyze environments and undertake actions aimed at achieving specific objectives, thus closely mirroring but not yet matching human cognitive functions.[4]

Central to AI's capability is machine learning (ML), a domain wherein algorithms analyze data, learn from it, and subsequently apply the acquired knowledge to make informed decisions. This process involves constructing mathematical models that facilitate predictions or decisions autonomously, without direct programming for specific tasks. A more specialized branch of ML, known as deep learning, utilizes layered neural networks that emulate human decision-making processes. These networks, capable of processing vast datasets, learn to recognize complex patterns and are crucial for tasks like image and speech recognition, as well as natural language processing.

Neural networks, essential components of deep learning, are organized into multiple layers: the input layer, one or more hidden layers, and an output layer. Each layer is composed of units or neurons that process incoming data from the previous layer before passing it to the next, allowing the system to learn increasingly complex features as the data progresses through the network.

Other AI Methodologies include rule-based modeling, data mining, fuzzy logic, case-based reasoning, and text and visual analytics.[5] Rule-Based Modeling uses if-then rules for knowledge representation, enabling the system to make decisions based on predefined logical rules. Data Mining involves extracting patterns from large datasets, which is pivotal for discovering hidden patterns and unknown correlations. Fuzzy Logic allows

---

[4] PHILIP BOUCHER, ARTIFICIAL INTELLIGENCE: HOW DOES IT WORK, WHY DOES IT MATTER, AND WHAT CAN WE DO ABOUT IT? VI (2020), http://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU(2020)641547_EN.pdf.

[5] Iqbal H. Sarker, *AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems*, 3 SN COMPUT. SCI. 1, 3–4 (2022).

for reasoning that is approximate rather than fixed and exact, handling the concept of partial truth—where values may range between completely true and completely false. Case-Based Reasoning solves new problems based on the solutions of similar past problems. Text and Visual Analytics involves extracting useful information from text and visual data to facilitate decision-making and insight generation.

The synthesis of these diverse AI methodologies typically involves combinatorial search and formal mathematical models, which are integral to developing systems capable of automated reasoning. This functionality is critical for AI's ability to process information and derive logical conclusions from intricate datasets.[6]

However, a significant challenge within AI, particularly evident in deep learning models, is their "black box" nature. The internal mechanisms by which these models make decisions are not always transparent, complicating efforts to understand and trace how inputs are transformed into outputs. This opacity can obstruct the debugging process, obscure bias detection and mitigation, and hinder comprehension of AI decision-making.[7] For example, Nvidia's self-driving cars learn from human behavior but might confuse the moon for a traffic light, and the DeepPatient project accurately predicted disease onset from medical records without providing explanations for its predictions, posing challenges for medical professionals trying to trust and interpret these AI systems.[8] This opacity not only makes it difficult to correct biases and errors but also potentially erodes trust, which is critical as AI becomes more embedded in everyday life. Each application benefits from AI's ability to swiftly process large datasets, learn from inputs, and enhance decision accuracy over time. Nevertheless, the deployment of AI systems brings forth critical ethical and practical considerations: including trust, privacy, and control. These aspects require ongoing evaluation and management to ensure that AI technologies are utilized responsibly and continue to align with human values and societal norms.[9]

AI systems continuously evolve by harnessing vast datasets and utilizing sophisticated learning algorithms to enhance their decision-making capabilities over time. This evolution is significantly supported by rapid advancements in hardware, which provide the necessary computational

---

[6] Stuart Russell, *The History and Future of AI*, 37 OXFORD REV. ECON. POL'Y 509, 512 (2021).
[7] Melissa Heikkila, *Nobody Knows How AI Works*, MIT TECH. REV. (Mar. 5, 2023), https://www.technologyreview.com/2024/03/05/1089449/nobody-knows-how-ai-works/.
[8] Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/?gad_source=1.
[9] *Id.*

power to process and analyze complex datasets efficiently. Moreover, the development of innovative architectures for structuring AI systems is essential, as these frameworks facilitate the effective interconnection and scaling of data layers, model complexity, and application breadth. Such structural innovations are crucial in optimizing the flow and processing of information, as well as enabling AI systems to handle more complex tasks, to adapt to new environments, and to integrate seamlessly across diverse applications. This dynamic interplay of data accumulation, algorithmic learning, hardware advancements, and system architecture refinements drives the continuous improvement and expansion capabilities of AI technologies.[10]

The primary hardware advancement critical to AI advancement has been increased computing power and chip performance. As AI algorithms become more complex and data-intensive, the demand for computational resources continues to escalate. Web3 systems offer a solution by enabling decentralized sharing of AI computing power. Through blockchain technology and distributed networks, unused computing resources can be harnessed and allocated efficiently, allowing AI tasks to be processed faster and at a lower cost. By democratizing access to computational resources, Web3 systems facilitate collaborative efforts in AI research and development, accelerating innovation and driving forward the capabilities of AI.[11]

The improvement in chip performance, particularly through GPUs and specialized AI processors, has been crucial in providing the high-speed, efficient computing required to train these models on large datasets. This combination of advanced AI models and robust computing infrastructure allows AI systems to process vast amounts of information rapidly, making strides toward achieving and even surpassing human-level thought in specific tasks. These technological enhancements not only increase the speed and efficiency of AI systems but also expand their applicability across different fields, pushing the boundaries of what AI can achieve and transforming potential theoretical concepts into practical applications.[12]

---

[10] GANESH D. GOVINDWAR & SHEETAL S. DHANDE, A REVIEW ON FEDERATED LEARNING APPROACH IN ARTIFICIAL INTELLIGENCE 2 (2022), https://doi.org/10.1109/ICCUBEA54992.2022.10010798.

[11] Andreas Sudmann, *The Democratization of Artificial Intelligence: Net Politics in the Era of Learning Algorithms*, 1 AI CRITIQUE 9, (Anna Tuschling, Andreas Sudmann & Bernhard J. Dotzler eds., 2019); IAN GOODFELLOW, YOSHUA BENGIO & AARON COURVILLE, DEEP LEARNING 11 (2016); Khaled Salah et. al, *Blockchain for AI: Review and Open Research Challenges*, 7 IEEE ACCESS 10127, 67 (2019), https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8598784.

[12] Jie Wen et al., *A Survey on Federated Learning: Challenges and Applications*, 14 INT'L J. MACH. LEARNING & CYBERNETICS 513, 513 (2023).

The dominance of big tech companies in AI development is significantly shaped by their comprehensive resources, which include access to vast datasets, state-of-the-art hardware, and extensive human capital. These advantages enable them to develop more advanced and sophisticated AI systems, providing big tech with a substantial competitive edge in the market. As a result, the solutions offered by these companies are not only more prevalent in the marketplace but also superior in technological capabilities. This disparity in resource availability and resultant AI proficiency creates a pronounced digital divide, where smaller companies and traditional sectors that lack similar access to large-scale data and cutting-edge technology struggle to keep pace. Consequently, this divide exacerbates the gap in AI capabilities, potentially leading to a concentration of power and influence in the hands of a few large tech entities, while others lag significantly behind in leveraging AI for innovation and improvement.[13]

Regulatory frameworks and ethical guidelines have become pivotal in shaping the evolution of AI, steering development towards a human-centric approach. These guidelines emphasize the creation of trustworthy AI systems that uphold privacy, ensure transparency, and strive for explainability. By prioritizing these principles, regulators aim to mitigate the risks associated with AI technologies, such as data misuse, biased decision-making, and opaque systems whose workings are not understandable to users or their developers. The push towards explainability is particularly crucial as it involves designing AI systems whose actions can be easily understood by humans, thus fostering greater trust and acceptance. This regulatory focus not only ensures that AI development aligns with ethical norms and societal values but also helps maintain public confidence in how these advanced technologies are integrated into daily life, making AI systems more accessible and safer for widespread use.[14]

The evolution of AI is increasingly guided by institutional regulations and ethical guidelines that advocate for a human-centric approach and the development of trustworthy AI systems that prioritize privacy, data governance, and transparency. However, strict data privacy regulations such as the General Data Protection Regulation (GDPR) present challenges to this growth trajectory. GDPR imposes stringent conditions on data sharing, which can limit the amount and variety of data AI systems use, potentially reducing their performance and exacerbating biases due to the restricted dataset. Furthermore, while the move towards more explainable, private, and transparent AI is commendable, these regulations can paradoxically consolidate power within large tech companies. These entities possess the

---

[13] Yannick Bammens & Paul Hünermund, *Using Federated Machine Learning to Overcome the AI Scale Disadvantage*, 65 MIT SLOAN MGMT. REV. 54, 55 (2023).

[14] JOSE LUIS ET AL., FED-XAI: FEDERATED LEARNING OF EXPLAINABLE ARTIFICIAL INTELLIGENCE MODELS 3–4 (2022), https://ceur-ws.org/Vol-3277/paper8.pdf.

vast data resources and sophisticated infrastructures needed to comply with such regulations and still develop effective AI solutions, unlike smaller companies or individual developers who might be sidelined by these legal constraints, thus altering the competitive landscape in favor of more established players.[15]

The future of AI's evolution is intricately linked to the widespread adoption of smartphones and the move towards decentralized data processing and model training. With smartphone usage projected to rise from over 5 billion users in 2020 to an estimated 7.5 billion by 2025, these devices are poised to become an even more significant foundation for AI development. The ubiquity of smartphones, coupled with their advanced computing capabilities and constant connectivity, offers a vast resource for deploying AI applications directly into the hands of users globally. This scale of interconnectivity and integration into daily life makes smartphones a crucial platform for AI. Furthermore, the evolution of AI technologies is increasingly embracing approaches like embedded and federated machine learning, which prioritize decentralization. These methods shift away from traditional cloud-centric models, opting instead to process data and train models directly on local devices. This approach not only leverages the computational power of individual smartphones but also enhances privacy and data security by keeping sensitive information on the device, thus shaping a future where AI applications are both powerful and privacy-preserving.[16]

AI Models

AI models have undergone a significant evolution from their early days of relying on symbolic reasoning and fuzzy logic, which depended heavily on predefined rule systems, to the modern paradigms that utilize ML and data mining. This shift marks a transition from rule-based to data-driven approaches, where current AI models excel at generating content and making decisions based on patterns and insights extracted from large, complex datasets. Unlike the earlier systems that operated within the strict confines of manually programmed rules and logical frameworks, these newer models adapt and learn from the data itself, leading to more dynamic, flexible, and scalable AI solutions. This ability to derive emergent systems

[15] Ye Yuan et al., *DeceFL: A Principled Fully Decentralized Federated Learning Framework*, 2 NAT'L SCI. OPEN 1, 2 (2023).
[16] Somdip Dey, *Are Embedded and Federated Machine Learning the Future of the AI Industry?*, FORBES (July 3, 2023), https://www.forbes.com/sites/forbestechcouncil/2023/07/03/are-embedded-and-federated-machine-learning-the-future-of-the-ai-industry/?sh=64bacd24515f.

from data allows for more robust and nuanced responses to a variety of tasks and challenges in real-time environments.[17]

AI systems are designed to interpret data, learn from this data, and utilize the acquired knowledge to accomplish specific objectives via flexible adaptation. These systems are typically classified into three types based on their underlying technologies and capabilities. First, analytical AI, which includes methods like fuzzy logic and symbolic reasoning, relies on predefined rules to process data. Second, human-inspired AI utilizes neural networks to mimic human brain functions, enabling it to learn from complex data sets and improve over time. Lastly, humanized AI extends capabilities to understand and replicate human emotions and social contexts, aspiring towards more comprehensive cognitive abilities. AI systems are also categorized by their evolutionary stages, which reflect their scope and sophistication: Artificial Narrow Intelligence (ANI) which excels in specific tasks, Artificial General Intelligence (AGI) that equals human cognitive abilities across a broad range of activities, and Artificial Superintelligence (ASI) which surpasses human intelligence and capability. Each stage represents an expansion in the scope, complexity, and autonomy of AI applications, moving from task-specific implementations to potentially self-aware systems that can perform across multiple domains with emotional and social intelligence.[18]

AI can be divided into four categories based on its potential functionalities, each representing different levels of complexity and capability. Currently, two of these categories are operational: (1) reactive machine AI, which responds to specific stimuli without past references, and (2) limited memory AI, which uses data from the recent past to make decisions. The other two categories, Theory of Mind AI and Self-Aware AI, remain theoretical and represent future aspirations in AI development. Theory of Mind AI aims to comprehend and simulate human emotional and cognitive processes, while Self-Aware AI seeks an even deeper understanding, potentially achieving consciousness. Developments like Emotion AI, which attempts to read and respond to human emotions, are steps towards these advanced, hypothetical stages of AI functionality.[19]

When a model incorporates machine learning, it typically utilizes one of three main types: supervised, unsupervised, or reinforcement learning, with semi-supervised learning acting as a hybrid between supervised and unsupervised. Supervised learning involves explicitly training the model by providing it with input and expected output data, which helps the system to

---

[17] BOUCHER, *supra* note 5, at 5–10.
[18] Michael Haenlein & Andreas Kaplan, *A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence*, 61 CAL. MGMT. REV. 4, 2–8 (2019).
[19] *Understanding the Different Types of Artificial Intelligence*, IBM (Oct. 12, 2023), https://www.ibm.com/blog/understanding-the-different-types-of-artificial-intelligence/.

learn and predict future outputs based on new inputs. Unsupervised learning, on the other hand, does not involve direct instruction; instead, the model analyzes unlabeled data to discover patterns and correlations on its own. Reinforcement learning is based on a behaviorist approach where the model learns to achieve a specific goal in a complex environment through trial and error, continuously adjusting its actions based on feedback to optimize the path towards the goal. This framework allows models to adapt and improve autonomously within their operational parameters and environments.[20]

Federated Model

In Federated AI Learning (FL), a comprehensive global model is created by aggregating individually trained local models, minimizing the need for direct data transfer. Unlike traditional centralized models, FL trains AI algorithms across multiple decentralized devices using local data samples, without requiring the data to be sent to a central server. This process involves each participating device (client) training an AI model on its own data locally, which ensures that sensitive information does not leave the device. These local models then communicate their parameters to a central server, which aggregates these parameters to update the global model. This method of parameter interaction, rather than direct data sharing, not only enhances privacy by keeping the data on the device but also facilitates a more efficient use of bandwidth. The updated global model is then sent back to each device, completing a cycle of learning that progressively improves the model's accuracy while maintaining data security.[21] This makes FL an effective solution for scenarios where privacy is paramount and data transmission costs are a concern.[22]

Federated Machine Learning (FedML) is a privacy-preserving collaborative AI approach that utilizes decentralized data, enabling entities with smaller datasets to leverage AI effectively. By distributing the computational tasks of ML across numerous devices that hold their data locally, FedML allows multiple participants to contribute to the creation of a robust model without needing to share their data. This method is particularly advantageous for smaller entities or those with privacy concerns, as it bypasses the necessity of large, centralized datasets that traditional ML models require. Furthermore, FedML's structure not only maintains data privacy but also facilitates the pooling of diverse data sources, enhancing the model's overall learning and predictive power. This approach aligns well

[20] Sara Brown, *Machine Learning, Explained*, MIT SLOAN (Apr. 21, 2021), https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.
[21] LUIS ET AL., *supra* note 15, at 4–5.
[22] GOVINDWAR & DHANDE, *supra* note 11, at 3–4; Jie Wen et al., *supra* note 13 at 517–518.

with small-data strategies, offering a practical solution for harnessing AI capabilities in scenarios where data availability is limited.[23]

In a FedML setup, the unique feature is that the ML model is trained across multiple decentralized servers, each using its distinct dataset. This framework enables smaller companies to participate in training an AI model collaboratively while retaining exclusive control over their data. Such a decentralized approach not only ensures the privacy and confidentiality of personal data but also complements other methods suited for small datasets, like transfer learning and self-supervised learning. By integrating these techniques, FedML facilitates the creation of community-trained large language models (LLMs), allowing even entities with limited data resources to contribute to and benefit from advanced AI technologies without compromising data security.[24]

Despite its advantages, practical implementations of FL encounter several bottlenecks that can affect model performance and efficiency, including significant privacy and security risks. These challenges arise because, while FL keeps data decentralized, it still involves the exchange of model parameters, which could potentially expose sensitive information if intercepted or improperly handled. Additionally, FL lacks theoretical guarantees that ensure reliability and robustness, making it less predictable for practical applications. Moreover, the rigid communication topology in FL, which typically requires constant coordination between numerous nodes, can lead to inefficiencies and does not easily adapt to dynamic network conditions or node failures. This rigidity can hinder the scalability and responsiveness of FL systems, complicating their deployment in environments with fluctuating data or network structures.[25]

FL can experience reduced accuracy due to the diverse and decentralized nature of its data sources, but this can be mitigated through various strategies. The decentralized setup in FL means that data variability across different nodes might lead to inconsistencies in training, which in turn can affect the overall FL model accuracy. To counter this, techniques such as data augmentation can enhance the volume and variety of data available for training, thereby improving the model's ability to generalize across different scenarios. Model compression techniques can optimize the processing capabilities of local devices, ensuring that even with limited hardware resources, the performance of the AI system is not compromised. Additionally, incentivizing participation by offering rewards or benefits for sharing higher-quality data or for more active involvement in the training process can significantly enhance the quantity and quality of data collected, thus boosting the overall effectiveness and accuracy of FL models. These

[23] Bammens & Hunermund, *supra* note 14, at 57.
[24] *Id.*
[25] GOVINDWAR & DHANDE, *supra* note 11, at 2.

strategies collectively aim to harness the full potential of decentralized learning while addressing the inherent challenges of data diversity.[26]

Centralized Model

In contrast to FL, centralized AI models necessitate the collection of data at a single, central location for processing, which is a common approach among major AI platforms. Traditional centralized ML approaches encounter significant obstacles when data are distributed across various locations and cannot be centralized due to privacy concerns or logistical challenges. The centralized model is employed by well-known entities such as Google, OpenAI, and Anthropic, where vast amounts of data are aggregated from various sources and then processed in a centralized server or data center. This approach allows for powerful computation and the application of complex algorithms across large datasets, potentially leading to more robust and sophisticated AI systems. However, it also raises concerns regarding privacy and data security, as the central accumulation of data increases the risk of data breaches and misuse. Centralized models, while efficient in terms of computational power and algorithmic sophistication, thus involve trade-offs that must be carefully managed, especially in terms of data governance and security protocols.[27]

Traditional centralized models necessitate transferring data to a central location for processing, whereas decentralized models enhance privacy but demand complex coordination to ensure consistent performance. Centralized systems, by pooling data from various sources into a single repository, can leverage powerful computational resources and advanced algorithms to optimize AI performance. However, this centralization can pose significant privacy risks and create a single point of failure. On the other hand, decentralized models, such as those used in FL, keep data localized on users' devices, significantly boosting data security and privacy. Yet, these models face challenges in synchronizing and coordinating between diverse and geographically dispersed devices to maintain uniform AI effectiveness. This coordination often requires sophisticated communication protocols and algorithms to manage and harmonize the distributed learning process, ensuring that performance does not vary widely across the network.[28]

To address these issues, novel methods like FL allow for the collaborative training of ML models without requiring the direct sharing of data. As discussed earlier, FL addresses these issues by decentralizing the learning process; algorithms are sent to local datasets where they are trained independently. The local models then send their learned parameters or

[26] Dey, *supra* note 17.
[27] GOVINDWAR & DHANDE, *supra* note 11, at 1; Jie Wen et al., *supra* note 13, at 514.
[28] BOUCHER, *supra* note 5, at 29–35.

updates back to a central server that aggregates these contributions into a global model. This method effectively sidesteps the need to compromise data privacy or tackle the logistical difficulties of data aggregation, making it ideal for scenarios where data sensitivity or distribution is a concern.[29]

In contrast to both centralized and FL models, embedded ML stands out by enabling devices to learn directly on their hardware, optimizing performance and enhancing privacy. This approach integrates ML algorithms directly into device firmware, allowing them to process data and learn from it locally without relying on cloud connectivity. By processing data on the device itself, embedded ML significantly reduces the need for constant data transmission to the cloud, which decreases latency and conserves bandwidth. Additionally, this method enhances user privacy by keeping data physically secured on the device. Despite these advantages, embedded ML faces challenges related to the finite computing resources available on the devices, which can constrain the complexity and capability of the AI models that can be implemented effectively.[30]

Embedded devices often face constraints due to limited resources, making it difficult to train complex models directly on the devices. To address these challenges, two primary strategies are commonly employed. First, developers might opt for simpler, less resource-intensive models that can operate effectively within the hardware limitations of embedded systems. These models require fewer computational resources, allowing them to run smoothly on devices with restricted processing power and memory. Second, transfer learning can be utilized to enhance the capabilities of embedded ML systems. This technique involves pre-training a model on a more powerful system with abundant data and then transferring the learned features to the embedded device, which only needs to fine-tune the model based on its specific application. This approach significantly reduces the computational burden on the embedded device while still leveraging advanced ML capabilities.[31]

2. AI GOVERNANCE

Current efforts to regulate AI applications have become increasingly evident across various jurisdictions in the United States. In 2023, legislative initiatives aimed at regulating AI applications were observed in at least twenty-five states, as well as Puerto Rico and the District of Columbia. These proposed bills encompass a wide range of uses and concerns, addressing both government and private sector applications of AI. They include stipulations for impact assessments to evaluate the potential effects

[29] LUIS ET AL., *supra* note 15, at 7–10.
[30] Dey, *supra* note 17.
[31] *Id.*

of AI systems, notification requirements to ensure transparency, guidelines for responsible use to maintain ethical standards, specific regulations for AI in healthcare to safeguard patient interests, and mandates for studies to continually assess and update AI regulations. These legislative actions reflect a growing recognition of the need to manage the complex impacts of AI technologies on society.[32]

To effectively implement AI governance, a multifaceted approach involving various tools and solutions is essential to address key issues. This approach should include promoting ethical frameworks, conducting rigorous research into the implications of AI, and developing measures for interpretability and explainability. Establishing norms, ethics, and values frameworks provides a foundation for guiding AI development and usage in a manner consistent with societal values. Additionally, researching the effects and implications of AI use is crucial for understanding its impact and identifying potential risks and benefits. These insights then inform the creation of technical solutions and legislative measures, which are necessary to address and mitigate the challenges posed by AI technology. Such comprehensive governance strategies ensure that AI systems are developed and deployed in ways that are beneficial, ethical, and transparent, thereby maximizing their positive impact while minimizing potential harms.[33]

AI governance fundamentally centers on ensuring accountability within AI systems, highlighting the obligation to comply with established guidelines and regulations throughout the AI lifecycle. Accountability in AI is multifaceted and can be evaluated through several key dimensions: the specific context in which AI is applied, which helps define its purpose and scope; the various stages of its lifecycle, including design, development, and deployment; the identification of all parties involved in these stages; and the stakeholders affected by AI's operations. Additionally, the standards used to judge AI—whether legal, ethical, or technical—play a crucial role. The mechanisms through which accountability is enforced, such as audits or compliance checks, are critical in ensuring that AI behaves as intended. Lastly, the consequences of applying the accountability framework determine the effectiveness of governance measures, influencing future policies and the development of AI technologies. This comprehensive approach to AI governance ensures that AI systems are not only effective and efficient but also operate within the bounds of societal norms and

[32] *Artificial Intelligence 2023 Legislation*, NAT'L CONF. OF STATE LEGISLATURES (Jan. 12, 2024), https://www.ncsl.org/technology-and-communication/artificial-intelligence-2023-legislation.

[33] James Butcher & Irakli Beridze, *What is the State of Artificial Intelligence Governance Globally?*, 164 RUSI J. 88, 89–90 (2019), https://doi.org/10.1080/03071847.2019.1694260.

regulations, maintaining public trust and safeguarding against potential misuses.[34]

The EU AI Act[35] exemplifies a risk-based legislative framework that categorizes AI applications according to the level of risk they pose, from "Unacceptable Risk" to "Transparency Requirements." This structure allows for tailored regulatory responses where the obligations for AI providers and users vary based on the identified risk level. AI systems that pose an "Unacceptable Risk" and are deemed a threat to people's safety or rights will be prohibited, although certain exceptions exist for law enforcement purposes. On the other hand, generative AI systems, while not classified as high-risk, are required to meet specific transparency obligations and comply with EU copyright laws. This approach aims to balance the benefits of AI innovation against potential harms, ensuring that AI technologies are used safely and ethically within the EU.[36]

AI regulation currently lacks global uniformity, with various regions adopting distinct strategies reflecting their unique priorities and values. The European Union adopts a precautionary principle, implementing bans on certain AI uses perceived as threats to rights or safety, emphasizing a controlled and safety-first approach. In contrast, the U.S. favors a more laissez-faire attitude, focusing on establishing guiding principles rather than stringent laws, thereby prioritizing innovation and the development of AI technologies without heavy governmental interference. Meanwhile, China presents a dual approach that aims to stimulate technological innovation while maintaining tight governmental control over AI applications, ensuring that advancements align with national interests and regulatory standards.

The divergence in regulatory philosophies on AI development and management reflects differing cultural, political, and economic priorities globally, despite the imperative for a unified approach to such transformative technology. The varying priorities of nations have led to distinct regulatory frameworks, shaped by cultural values, political ideologies, and economic interests. However, achieving a consensus on AI regulation is crucial to address ethical, societal, and technological challenges effectively. Without global cooperation and alignment, disparities in AI

---

[34] Claudio Novelli, Mariarosaria Taddeo & Luciano Floridi, *Accountability in Artificial Intelligence: What It Is and How It Works*, AI & SOC'Y (2023), https://doi.org/10.1007/s00146-023-01635-y.

[35] *Shaping Europe's Digital Future*, EUROPEAN COMM'N (May 6, 2024), https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

[36] EU AI act: First regulation on artificial intelligence , TOPICS | EUROPEAN PARLIAMENT, https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (last visited June 16, 2024) [hereinafter *EU AI Act*].

governance may hinder innovation, exacerbate inequalities, and impede the realization of AI's full potential for the benefit of humanity.[37]

Shortcomings in Existing AI Governance

AI governance does encounter a critical challenge in mitigating biases within AI systems, where biases can inadvertently arise through algorithms incorporating discriminatory practices due to data used in training. These biases generated manifest in various forms, such as skewed considerations for maternity leave or inadequate recognition of foreign qualifications, which can perpetuate inequality. A specific example of this can be seen with incidents like "Tay, the racist chatbot,"[38] which reflected and amplified social prejudices due to flawed training data. As such, these issues demonstrate a profound misalignment between AI operations and societal values, ethics, and norms.[39]

The unpredictable nature and potential biases of AI models call for a cautious and research-based governance approach. As AI technologies evolve, establishing a governance framework for prioritizing transparency and realistic expectations becomes critical. Such a framework should focus on clear communication about what AI can and cannot do, correcting the often overly optimistic or misleading portrayals presented by tech industry marketing. Additionally, it should aim to anchor public discourse and policy-making in a realistic understanding of AI's capabilities and limitations, thereby ensuring that the development and deployment of AI technologies are aligned with ethical standards and societal needs. This approach will help mitigate risks and guide the responsible integration of AI into various sectors.[40]

Legal and ethical challenges are heightened when AI is deployed in critical decision-making roles that significantly impact human lives, particularly when the reasoning behind AI's decisions is opaque. AI algorithms are increasingly used to make determinative decisions about parole eligibility, employment, and medical strategies—areas where the consequences for individuals are profound.[41] The reliance on purely quantitative data in these decisions raises substantial ethical concerns about the sidelining of human empathy and qualitative judgment. Furthermore, the growing ubiquity of Generative AI systems in such roles exacerbates these issues. These systems, while enhancing operational efficiency, obscure the

[37] Matthew Hutson, *Rules to Keep AI in Check: Nations Carve Different Paths for Tech Regulation*, NATURE (Aug. 8, 2023), https://www.nature.com/articles/d41586-023-02491-y.
[38] BOUCHER, *supra* note 5, at 22.
[39] *Id.* at 22–26.
[40] Heikkila, *supra* note 8.
[41] Knight, *supra* note 9.

decision-making process further, making it difficult to discern how conclusions are reached. This opacity not only challenges the principles of fairness and accountability but also risks diminishing the human elements essential in sensitive contexts, thus intensifying the ethical dilemmas associated with AI governance.[42]

The ethical landscapes and privacy concerns surrounding AI underscore the critical need for robust guidelines to prevent misuse and ensure that AI technologies serve the greater good while safeguarding individual privacy rights. As AI becomes increasingly integrated into diverse sectors—from healthcare to finance—it becomes imperative to enforce regulatory compliance tailored to each sector's specific needs and risks. This regulatory framework must address the dual challenge of harnessing AI's potential for societal benefits while effectively managing the risks associated with data privacy and ethical dilemmas. Establishing these guidelines involves a collaborative effort among technologists, legal experts, policymakers, and the public to create a balanced approach that promotes innovation while protecting fundamental human rights.[43]

Privacy concerns in AI governance are heightening worries about how personal data may be abused or manipulated. As AI technologies evolve, they rely heavily on vast amounts of data, transforming personal information from a resource that individuals could control and use at their discretion into a fundamental operational tool for AI systems. This shift not only increases the potential for misuse of personal data but also complicates individuals' ability to understand and manage how their information is utilized. Consequently, the depth and breadth of data required for AI operations raise significant concerns about privacy, as these systems can potentially exploit personal data in ways that are difficult to predict and often beyond the direct control of the individuals whose data is being used.[44]

The debate around AGI amplifies existing concerns regarding AI's misuse, particularly focusing on the profound existential risks and long-term social impacts. The advent of AGI could radically transform job markets through widespread automation, potentially rendering many career fields redundant. This scenario indicates that no sector may be safe from disruption. Such a significant shift highlights the critical necessity for balanced regulations that encompass not only the transparency and safety of AI models but also tackle their extensive economic and existential effects. Crafting these regulations is a delicate task; it is imperative to strike a balance that encourages technological innovation while ensuring that the

---

[42] *Id.*

[43] *What is Artificial Intelligence?*, IBM (Mar. 4, 2024), https://www.ibm.com/topics/artificial-intelligence.

[44] Francois Candelon et al., *AI Regulation Is Coming: How to Prepare for the Inevitable,* HARVARD BUS. REV. (2021), https://hbr.org/2021/09/ai-regulation-is-coming.

deployment of AI does not negatively impact societal structures or economic stability, thereby safeguarding against disruptive consequences across all levels of employment and industry sectors.[45]

The notable cases involving Apple and Amazon have intensified scrutiny over biased outcomes produced by AI systems. In specific incidents, Apple's credit card algorithms and Amazon's resume scanning technology were accused of exhibiting bias, which led to widespread criticism and brought to light the potential discriminatory practices embedded within AI applications. These instances exemplify how even well-intentioned AI systems, developed by leading technology companies, can inadvertently perpetuate inequalities. The fallout from these cases highlights the critical need for rigorous testing, transparency in AI decision-making processes, and mechanisms for addressing any biases that AI systems may harbor. It also highlights the importance of continuous monitoring and adaptation of AI systems to prevent discriminatory outcomes and ensure fairness across all user interactions.[46]

Decentralizing AI Governance

Decentralized ML governance provides a robust alternative to centralized systems by bolstering data protection and minimizing vulnerabilities. This approach significantly enhances the management of digital identities within ML ecosystems, promoting the adoption of self-sovereign identities, soul-bound tokens, and decentralized identifiers. Self-sovereign identities allow individuals to manage their personal data independently, without reliance on third-party storage, thereby ensuring privacy and control. Soul-bound tokens, which are non-transferable and linked to specific digital identities, further secure user data by preventing unauthorized transfers. Additionally, decentralized identifiers, standardized by the World Wide Web Consortium (W3C), offer a reliable method for unique identification on the internet without depending on centralized registries. Together, these technologies empower individuals with unprecedented control over their digital identities, making ML systems more secure and user-centric.[47]

Platforms like SingularityNET harness distributed ledger technology (DLT) to enhance interactions between AI services and users, promoting

---

[45] Bill Whyman, *AI Regulation is Coming: What is the Likely Outcome?*, CTR. FOR STRATEGIC & INT'L STUDIES (Oct. 10, 2023), https://www.csis.org/blogs/strategic-technologies-blog/ai-regulation-coming-what-likely-outcome.

[46] *Id.*

[47] Dana Alsagheer, Lei Xu & Weidong Shi, *Decentralized Machine Learning Governance: Overview, Opportunities, and Challenges*, 11 IEEE ACCESS 96718, 6–7 (2023), https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10238468.

autonomy and privacy. By leveraging blockchain, SingularityNET allows AI developers to maintain control over their intellectual property while fostering the creation of decentralized datasets equipped with robust privacy controls. This setup facilitates direct transactions between AI services and agents without the need for intermediaries, thereby ensuring a decentralized and transparent operational environment. Such a platform not only streamlines the delivery of AI-driven services but also enhances the security and privacy of data exchanged, making it a vital tool for developers looking to deploy and manage AI solutions efficiently while safeguarding user and organizational data.[48]

The decentralized governance model facilitated by overlapping institutions presents a promising approach for effective AI regulation, characterized as a regime complex. This description reflects the current global AI governance landscape, which is notably fragmented yet adaptable, benefiting from a system that encompasses multiple, autonomously coordinated values. Such a structure promotes decentralization, potentially enhancing the effectiveness, inclusivity, and adaptability of AI governance. However, the apparent disorganization within this regime complex also underscores the urgent need for normative research focused on ethical values. This research would provide critical guidance for the development of a more unified and coherent governance framework, ensuring that AI development aligns with broader societal goals and ethical standards.[49]

A decentralized approach to AI governance acknowledges that AI's global externalities surpass national boundaries, requiring the development of international standards to balance innovation and risk mitigation. As AI technologies increasingly influence global systems, the actions of multinational corporations in driving cross-border AI development emphasize the need for a global cooperative regulatory framework. This unified approach aims to prevent the fragmentation of international policies and ensure cohesive practices across borders. Significant issues such as accountability, the risk of dehumanization, and the competitive dynamics surrounding military AI development further highlight the critical need for these discussions. Consequently, global governance forums are becoming increasingly vital in establishing joint regulations that address these complex

[48] Gabriel Axel Montes & Ben Goertzel, *Distributed, Decentralized, and Democratized Artificial Intelligence*, 141 TECH. FORECASTING & SOC. CHANGE 354, 354–58 (2019), https://doi.org/10.1016/j.techfore.2018.11.010.

[49] Jonas Tallberg et al., *The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research*, 25 INT'L STUD. REV. 1, 8 (2023), https://academic.oup.com/isr/article/25/3/viad040/7259354.

challenges, ensuring that AI development remains safe, ethical, and beneficial worldwide.[50]

Furthermore, a decentralized approach to AI governance could shift the balance of power, enhancing user autonomy and negotiation capabilities. By introducing the concept of "computational agency," this approach empowers end-users to control their own data and actively participate in service agreements, challenging the current norm of accepting terms without genuine negotiation. Typically, users are presented with standard terms that they can either accept or decline without the opportunity for dialogue. The proposal for computational agency seeks to disrupt this dynamic, providing users with the tools necessary to negotiate terms effectively. This empowerment enables a higher degree of personalization and privacy, leading to more favorable service outcomes that truly meet user needs and protect their interests.[51]

Blockchain and DLT Technology

The fusion of blockchain and AI signifies a crucial advancement in digital governance, offering a framework for the transparent, secure, and accountable management of AI. This integration ensures that as AI continues to evolve and become integral to various sectors, it does so in a manner that is ethical, responsible, and aligned with societal values. The collaborative potential of blockchain and AI paves the way for a future where digital systems are governed with an unprecedented level of integrity and accountability, ensuring that technological advancements contribute positively to society. Blockchain's inherent characteristics such as decentralization, immutability, and transparency, complement AI's need for data integrity and auditability. This synergy can significantly enhance the oversight and control mechanisms for AI systems, making it easier to track and verify AI decisions and actions, thereby reducing the risks of misuse and enhancing trust among users and regulators alike.

The intersection of blockchain technology and AI represents a significant paradigm shift in digital governance, offering new methodologies for managing and ensuring the ethical application of AI. This integration heralds a transformative era where AI's sophisticated decision-making processes are complemented by blockchain's immutable record-keeping capabilities, promising enhanced accountability and transparency.

---

[50] Peter Cihon, Matthijs M. Maas & Luke Kemp, *Fragmentation and the Future: Investigating Architectures for International AI Governance*, 11 GLOB. POL'Y 545, 547 (2020), https://onlinelibrary.wiley.com/doi/epdf/10.1111/1758-5899.12890.

[51] WENJING CHU, A DECENTRALIZED APPROACH TOWARDS RESPONSIBLE AI IN SOCIAL ECOSYSTEMS 79 (2022), https://doi.org/10.48550/arXiv.2102.06362.

Blockchain as a Governance Mechanism for AI

Blockchain technology, with its distinct attributes of tamper-resistance, sequential data organization, and enhanced security, emerges as a formidable solution to AI's governance challenges. It provides a transparent and unalterable ledger of AI's data inputs and changes, facilitating traceability and accountability in AI operations.[52]

The capability of blockchain to offer an immutable log of transactions and modifications within AI systems is crucial for industries leveraging AI for decision-making. This feature allows stakeholders to track the lineage of AI decisions back to their original data inputs, enabling easier identification and correction of errors.

The integration of blockchain into AI democratizes the governance of AI by making its operations more accessible and understandable to a wider range of stakeholders. This transparency is crucial for building trust in AI systems and ensuring they are utilized responsibly and ethically.

Blockchain's governance capabilities ensure that AI systems not only operate more efficiently and autonomously but also adhere to ethical standards and regulatory requirements. This alignment is essential for fostering a digital ecosystem that is sustainable, trustworthy, and socially responsible. Blockchain's ledger is immutable, meaning once data is entered, it cannot be altered or deleted. This feature is crucial for AI governance, as it ensures a transparent and unchangeable history of AI decisions, data inputs, and modifications. Stakeholders can audit these records at any time to verify the AI's actions and the data it processed, enhancing trust and transparency. This transparency is fundamental in ensuring that AI systems are accountable and operate within ethical guidelines and legal frameworks. Blockchain's secure nature protects against unauthorized access and tampering, safeguarding the data AI systems use and generate. This security is essential for maintaining the privacy and integrity of sensitive information, aligning AI operations with privacy laws and ethical considerations. Moreover, blockchain can facilitate secure data sharing among AI systems, promoting collaboration while adhering to confidentiality requirements.

The traceability afforded by blockchain's ledger allows for tracking the provenance of data used by AI systems, ensuring that the data is accurate, lawful, and ethically sourced. Should an AI system make a decision that leads to a dispute or investigation, the blockchain can provide an auditable trail of all relevant actions and data inputs, establishing accountability and facilitating resolution in accordance with ethical and regulatory standards.

---

[52] SATOSHI NAKAMOTO, BITCOIN: A PEER-TO-PEER ELECTRONIC CASH SYSTEM 2–8, https://www.bitcoin.org/bitcoin.pdf (last visited May 31, 2024).

Blockchain operates on a decentralized model, where control and decision-making processes are distributed across a network rather than centralized in a single entity. This decentralization democratizes AI governance, allowing multiple stakeholders to participate in the oversight and decision-making processes. It prevents any single party from having undue influence over the AI, promoting a fair and balanced approach to governance that aligns with ethical standards and societal values.

Smart contracts are self-executing contracts with the terms of the agreement directly written into code. They can be used to automate compliance with regulatory requirements and ethical guidelines. For instance, smart contracts can automatically enforce privacy laws by controlling AI's access to personal data based on predefined rules. This automated compliance ensures that AI systems operate within legal boundaries, reducing the risk of violations and enhancing societal trust.

An argument can be made that the governance capabilities of blockchain and DLT technology can help ensure that AI systems not only enhance their operational efficiency and autonomy but also rigorously adhere to ethical standards and regulatory requirements. By embedding these principles into the foundation of AI systems, blockchain fosters a digital ecosystem that is sustainable, trustworthy, and socially responsible. This alignment is crucial for maximizing the benefits of AI while mitigating risks and ensuring that technological advancements contribute positively to society.

Adapting Decentralization of AI Governance to AI Models

While decentralization can offer benefits such as autonomy and local control, it also introduces challenges that require careful consideration and innovative solutions. Achieving effective AI governance in a federated model necessitates addressing these challenges through coordination, standardization efforts, collaborative frameworks, and mechanisms to ensure accountability and transparency.

Decentralized structures, while beneficial in many ways, often encounter challenges related to efficiency and operational barriers, particularly evident in the environmental agreement domain. These structures can complicate both negotiations and monitoring processes due to the lack of a central authority to streamline decisions and enforce agreements. In the context of environmental agreements, this decentralization can lead to difficulties in coordinating among multiple parties, each with their own interests and priorities, which may delay consensus and impede the swift implementation of necessary measures. Furthermore, monitoring compliance in such decentralized settings is challenging, as it requires effective collaboration and communication across diverse entities to ensure that all parties adhere to agreed-upon standards and

commitments. This can ultimately affect the overall effectiveness and responsiveness of environmental initiatives.[53]

Traditional ML relies on centralized data pipelines for model training, but the inherently fragmented nature of data presents significant challenges for collaboration and privacy. Sending decentralized datasets to a central server for processing raises serious privacy concerns and vulnerability issues, particularly if the central server fails or is attacked. In response, privacy-preserving frameworks like federated learning have been developed, yet these often still depend on a central client to collect and distribute model information, resulting in high communication loads and centralized vulnerabilities.

In the federated model of AI governance, many challenges cannot be easily decentralized. For example, different entities or organizations maintain their own AI systems and datasets. This can lead to variations in standards, protocols, and formats used, making it difficult to establish a unified governance framework. Standardization is crucial for interoperability, collaboration, and ensuring ethical practices, which all become challenging in a decentralized setting.

In a federated model, where decision-making power is distributed across multiple entities, it becomes harder to achieve consensus and cooperation, making it challenging to establish cohesive AI governance mechanisms. In a federated model, it can be challenging to ensure transparency and accountability across all participating entities due to the lack of centralized control, which this author does not otherwise advocate. This can lead to issues such as biased or unfair AI systems, inadequate privacy protection, or unequal access to AI benefits.

In a federated model, enforcing regulations and policies related to AI can be complex. In a decentralized setting, different entities may have varying interpretations of regulations or differing levels of commitment to compliance. This can hinder effective enforcement, monitoring, and oversight of AI systems, potentially leading to misuse or unethical practices.

A more decentralized web3 model of AI governance as proposed here could ostensibly address these challenges in a federated AI governance model by distributing governance more equitably across network participants.[54] That model could truly democratize AI governance, ensuring no single entity can dominate decision-making processes. This would align AI development and deployment with broader societal values and ethical standards.

---

[53] Peter Cihon, Matthijs M. Maas & Luke Kemp, *Fragmentation and the Future: Investigating Architectures for International AI Governance*, 11 GLOB. POL'Y 545, 550 (2020), https://onlinelibrary.wiley.com/doi/epdf/10.1111/1758-5899.12890.

[54] CRAIG CALCATERRA & WULF KAAL, DECENTRALIZATION: TECHNOLOGY'S IMPACT ON ORGANIZATIONAL AND SOCIETAL STRUCTURE 66, 62 (2021).

To address the issues with decentralizing the federated AI model, several intermediary solutions have been proposed in the literature. Among those is Decentralized Federated Learning (DeceFL). DeceFL represents an evolution of FL by addressing its central shortcomings, particularly by enabling direct local information exchange among clients without the need for centralized aggregation. In contrast to centralized models, traditional FL, and even swarm learning, DeceFL is more decentralized, meaning there is no central server collecting or processing data, which enhances privacy and system resilience. It also maintains competitive performance with no noticeable gap when compared to other models. All clients in DeceFL have equal access to the model, ensuring fairness in training and usage. Additionally, the system is adaptable to various network structures, performing effectively across different setups, and importantly, it avoids the communication of privacy-sensitive data among clients. These attributes not only reduce the communication load but also significantly bolster the privacy and applicability of the system, making DeceFL particularly suited for sectors like healthcare and smart manufacturing where data security and system resilience are paramount.[55]

The DeceFL approach ensures that every client can reach the global minimum with zero performance gap and achieve the same convergence rate as centralized methods when the loss function is smooth and strongly convex. The effectiveness of the DeceFL algorithm has been demonstrated across various applications, including those with convex and nonconvex loss functions, and in settings with time-invariant and time-varying network topologies, as well as IID and non-IID datasets. This showcases DeceFL's broad applicability and potential impact on real-world scenarios, particularly in medical and industrial fields.[56]

Decentralized AI governance solutions in the federated model are complicated by the need for compliance with stringent data privacy regulations such as the General Data Protection Regulation (GDPR). GDPR highlights the need for robust data management practices and effective solutions to auditing challenges. Decentralized systems, by their nature, distribute data across various nodes, making it difficult to ensure all data handling meets the strict privacy standards set by regulations like GDPR. This structure necessitates advanced data management strategies that can secure data across a dispersed network and complex auditing processes to verify compliance continuously. Addressing these challenges is crucial for maintaining the integrity of decentralized systems and ensuring they operate

[55] Yuan et al., *supra* note 16, at 6–7.
[56] *Id.*

within legal frameworks, thereby safeguarding user privacy and enhancing trust in these technologies.[57]

Managing ML assets and adhering to laws such as GDPR and CCPA is significantly more challenging under decentralized governance, raising concerns over privacy and data management. In decentralized environments, where data and operations are spread across various stakeholders and locations, ensuring consistent compliance with stringent privacy laws becomes complex. The distributed nature of these systems complicates the tracking of data flows and the enforcement of privacy controls, making it difficult to demonstrate compliance during audits. This challenge is exacerbated when multiple stakeholders are involved, as coordination and transparency across different entities must be maintained to ensure that all aspects of the system adhere to legal standards. Thus, decentralized governance demands robust mechanisms for compliance verification to address these inherent difficulties in managing privacy and data security effectively.[58]

III. PROPOSED SYSTEM

One promising way to address the decentralized governance needs of AI pertains to the implementation of Decentralized Autonomous Organizations (DAOs) for the governance of AI through expert community consensus. The integration of Decentralized Community Governance, blockchain technology, and federated communications platforms facilitates a robust mechanism for overseeing AI development and application. This system hinges on the interplay between various components including a Layer 1 blockchain, a communication forum platform, and specialized smart contracts.

The proposed model presents a sophisticated approach to decentralized governance, leveraging blockchain technology and federated communication platforms to create a dynamic, transparent, and participatory environment for managing and executing tasks. By intertwining the technological capabilities of different components with innovative governance mechanisms like validation pools and smart contracts, this system exemplifies the potential for decentralized autonomous organizations to enhance digital collaboration and decision-making for AI governance.

Foundations

[57] Dana Alsagheer, Lei Xu & Weidong Shi, *Decentralized Machine Learning Governance: Overview, Opportunities, and Challenges*, 11 IEEE ACCESS 96718, 13 (2023), https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10238468.
[58] *Id.*, at 12.

A DAO operates on a blockchain and is characterized by distributed governance mechanisms, enabling stakeholders to collectively make decisions without centralized authority.[59] Key terminology and functions of the proposed system facilitate the core mechanisms as outlined below.[60] Matrix serves as a federated communication platform, with Synapse being its reference server implementation. This setup provides a decentralized communication layer for DAO operations.[61]

The DAO operates a forum as an on-chain collection of posts, each uniquely identified and possibly citing previous contributions. This forum forms a Weighted Directed Acyclic Graph (WDAG), facilitating organized discussion and citation among participants.

Validation Pool (VP) form a crucial mechanism where an author's stakes are pooled to evaluate specific posts within the forum. The outcome of a VP can lead to the minting of new REP tokens, reflecting the consensus on a given issue or contribution.

Work Evidence and Work Smart Contract represent the tangible output of work that meets certain criteria. Work Smart Contracts manage the logistics of work contracts, staking, and the validation process, ensuring transparency and fairness in task assignment and completion.

---

[59] VITALIK BUTERIN, A NEXT-GENERATION SMART CONTRACT AND DECENTRALIZED APPLICATION PLATFORM 34, 4 (2014), https://ethereum.org/en/whitepaper/.

[60] Key Terminology and Components:

- Decentralized Autonomous Organization (DAO): A digital organization operated by smart contracts on a blockchain, facilitating decentralized governance without central authority.
- Community Governance Framework (CGF): The structural rules and protocols defining the operation and governance of a DAO.
- Matrix: A federated communications platform providing decentralized communication across various services.
- Synapse: The reference server implementation for Matrix, enabling federated communication.
- Element-web: The reference client (front-end) implementation for Matrix, facilitating user interaction with the Matrix ecosystem.
- Homeserver: A server running the Synapse software or an equivalent, serving as a node in the Matrix network. Functional Overview:
- Client Integration: An Element-web widget integrates with the Casper Wallet, allowing users to participate in the DAO directly from any Matrix room.
- Roles within the DAO:
    - Worker: A participant with a web3 Wallet who can contribute work, stake, and earn reputation (REP) and CSPR.
    - Customer: A participant who initiates work contracts by paying with CSPR.
- Reputation (REP): A non-fungible token (NFT) representing the credibility and contributions of a worker within the DAO, based on the outcomes of validation pools and forum citations.

[61] *About Matrix*, MATRIX, https://matrix.org/about/ (last visited June 6, 2024).

Membership in the DAO is signified by holding REP tokens, which grant voting rights and a share in DAO revenues. The dynamic valuation of REP tokens through validation pools allows for a flexible and responsive governance model that adapts to the collective decisions of the DAO members.

On top of the foundational governance structure, the DAO can implement various contracts for specific operational needs, such as work assignments and availability management. Given the cost of on-chain operations, a strategy involving off-chain activities consolidated into on-chain posts (roll-ups) creates efficiency.

The integration with Matrix ensures a decentralized repository and communication platform, enabling seamless interaction and data exchange among DAO participants.

Governance Mechanism

Decentralized Community Governance is central to this decentralized governance model, employing a combination of smart contracts, a reputation system (REP), and a validation pool mechanism to ensure AI governance aligns with expert community consensus.

Reputation (REP) Tokens are Non-Fungible Tokens (NFTs) represent an individual's contribution and standing within the DAO, allowing them to participate in governance decisions. The ERC 721 and ERC 1155 extension enables the association of a numeric value with each token, reflecting the individual's reputation.[62]

The Forum of the proposed governance system is an on-chain data structure of posts forming a weighted directed acyclic graph (WDAG), facilitating the documentation and citation of contributions. Validation Pools are consensus operations targeting specific forum posts, enabling the minting and distribution of REP based on community consensus.

By leveraging the Decentralized Community Governance within a DAO, the input parameters and learning data for AI systems can be governed through expert community consensus. This process involves submitting proposals to the Forum and undergoing Validation Pool review, ensuring that only vetted and consensus-backed data and parameters are utilized in AI development.

Work Smart Contracts and Availability Smart Contracts operationalize the transactional aspects of AI governance. Work Smart contracts define the terms under which AI governance tasks are undertaken, while Availability

[62] GAVIN WOOD, ETHEREUM: A SECURE DECENTRALISED GENERALISED TRANSACTION LEDGER 7 (2014), http://cryptochainuni.com/wp-content/uploads/Ethereum-A-Secure-Decentralised-Generalised-Transaction-Ledger-Yellow-Paper.pdf.

Smart Contracts facilitate the assignment of these tasks to reputable community members, ensuring accountability and quality in AI governance tasks.

This governance model promotes decentralization by distributing decision-making authority across a wide range of experts, rather than centralizing it in the hands of a few. Through the Validation Pool mechanism, expert consensus is required for significant decisions, ensuring that AI governance reflects the collective expertise and ethical considerations of the community.

Precedent and Citation System

The application of Weighted Directed Acyclic Graphs to the governance of AI through a precedent and citation system offers a robust, scalable, and dynamic framework suitable for managing the rapid evolution of AI technologies. By leveraging the structural advantages of WDAGs, such a system can ensure that AI governance remains effective, relevant, and responsive to the ever-changing landscape of AI development and its societal impacts. This approach not only facilitates the practical management of AI governance but also supports the ethical and legal compliance of AI systems, aligning them with societal values and regulatory requirements. This system's significance is amplified in the context of AI's exponential growth, necessitating governance mechanisms that can adapt and scale accordingly.

A WDAG is a fundamental construct in the field of graph theory and computer science, characterized by its directed edges, absence of cycles, and assignment of weights to each edge. This mathematical structure provides a powerful framework for representing relationships and processes that have inherent directionality, precedence constraints, and varying degrees of importance or capacity among their connections.

A WDAG consists of vertices (or nodes) connected by directed edges (or arcs), where each edge has an associated weight. The directed nature of the edges means that each connection between two vertices has a designated direction, indicating the flow from one vertex to another. The acyclic characteristic ensures that there are no loops within the graph, meaning it is impossible to start at a vertex and follow a sequence of directed edges that eventually loops back to the starting vertex.[63]

The weights assigned to the edges in a WDAG can represent various quantitative attributes such as cost, distance, time, or capacity, depending on the specific application. These weights play a crucial role in algorithms that

[63] JORGEN BANG-JENSEN & GREGORY Z. GUTIN, DIGRAPHS: THEORY, ALGORITHMS, AND APPLICATIONS 6 (2d ed. 2009), https://link.springer.com/content/pdf/10.1007/978-1-84800-998-1.pdf.

operate on WDAGs, influencing the computation of shortest paths, scheduling, and other optimization problems.

WDAGs find extensive applications across various domains including computer science, operations research, and engineering. One notable application is in task scheduling, where tasks are represented by vertices, and precedence relationships (i.e., the requirement that one task must be completed before another can begin) are represented by directed edges. The weights on these edges can indicate the time required to complete tasks or the transition time between tasks, aiding in the efficient scheduling of tasks to minimize overall completion time or resource utilization.[64]

In the context of network routing and communication, WDAGs can model network topologies where messages or data packets must be transmitted across a network without cycles, ensuring efficient data flow. The weights on the edges can represent bandwidth, latency, or other network characteristics, influencing routing decisions to optimize performance and resource allocation.[65]

Additionally, WDAGs are instrumental in project management and planning, particularly in the application of the Critical Path Method (CPM) for project scheduling. This method leverages WDAGs to model project tasks, their dependencies, and durations, identifying the longest path through the graph (the critical path) that determines the minimum project duration.[66]

The algorithmic treatment of WDAGs involves specialized algorithms for traversing the graph, computing shortest paths, and identifying topological orderings—a linear ordering of its vertices that respects the direction of the edges, which is particularly useful for scheduling and planning applications.[67]

Dynamic Governance

The structure of WDAGs, with their vertices representing legal precedents or governance rules and directed edges signifying citations or logical dependencies, creates an ideal model for organizing and navigating the multitude of governance considerations pertinent to AI. In such a system, the weights on the edges could quantify the relevance, authority, or impact of each precedent or citation, guiding the decision-making processes

[64] EDWARD G. COFFMAN, COMPUTER AND JOB-SHOP SCHEDULING THEORY 7 (1976).
[65] DIMITRI BERTSEKAS & ROBERT GALLAGER, DATA NETWORKS 120–125 (1987).
[66] James E. Kelley & Morgan R. Walker, *Critical-Path Planning and Scheduling*, PROCEEDINGS OF THE EASTERN JOINT COMPUTER CONFERENCE 160, 161–167 (1959), https://doi.org/10.1145/1460299.1460318.
[67] A. B. Kahn, *Topological Sorting of Large Networks*, 4 COMMC'NS ACM, 558, 559–562 (1962), https://dl.acm.org/doi/pdf/10.1145/368996.369025.

in AI governance by highlighting the most pertinent and influential governance frameworks or legal precedents.

The dynamic nature of WDAGs is essential for AI governance for several reasons, including but not limited to scalability and flexibility, navigating complexity, and decision support.

As AI technologies evolve, new precedents and governance rules will emerge. WDAGs can seamlessly integrate these new elements without disrupting the existing structure, ensuring the governance framework remains comprehensive and up-to-date.[68]

The complexity of AI systems and their potential impact across different sectors demands a nuanced approach to governance. WDAGs allow for the mapping of intricate relationships between governance rules and their applications, enabling stakeholders to navigate this complexity more effectively.[69]

By assigning weights to the edges based on factors such as recency, jurisdictional relevance, or cited frequency, a WDAG-based system can aid in prioritizing certain governance pathways over others, supporting more informed decision-making in AI governance.[70]

Exponential Evolution of AI and Governance Needs

The exponential evolution of AI underscores the necessity for governance frameworks that can not only keep pace with technological advancements but also anticipate future governance challenges. WDAGs, with their inherent adaptability and capacity for organizing complex information, provide a mechanism for such anticipatory governance.[71] This adaptability ensures that as AI systems learn and evolve, the governance mechanisms guiding their development and application can evolve correspondingly. Decentralized AI governance utilizing WDAGs also offers a comprehensive framework capable of addressing the complexities and rapid evolution of AI technologies. This governance approach, grounded in the structural benefits of WDAGs, ensures the governance mechanisms are scalable, dynamic, and aligned with ethical, legal, and societal expectations. By applying WDAGs to AI governance through a precedent and citation system, we can create a robust infrastructure that supports the ethical

[68] JORGEN BANG-JENSEN & GREGORY Z. GUTIN, DIGRAPHS: THEORY, ALGORITHMS, AND APPLICATIONS 6 (2d ed. 2009), https://link.springer.com/content/pdf/10.1007/978-1-84800-998-1.pdf.
[69] EDWARD G. COFFMAN, COMPUTER AND JOB-SHOP SCHEDULING THEORY 7 (1976).
[70] A. B. Kahn, *Topological Sorting of Large Networks*, 4 COMMC'NS ACM, 558, 559–562 (1962), https://dl.acm.org/doi/pdf/10.1145/368996.369025.
[71] FESTSCHRIFT, *supra* note 2.

development and application of AI, fostering trust and compliance within the AI ecosystem.

Dynamic Real-Time Governance

The above illustrated key system features and WDAG components contribute a much needed innovative approach to AI governance. The proposed system leverages the principles of decentralization and real-time data analytics to enhance dynamic AI governance. This approach is particularly relevant in the context of managing and governing AI systems, where ethical and legal standards are constantly evolving due to technological advancements and shifting societal values. Societal values are reflected in the core of the proposed system as the WDAG allows for dynamic governance where precedents in the system evolve much faster and dynamically in comparison with legacy systems.

The proposed WDAG system can harness decentralized networks to gather and assess community sentiment and ethical considerations in real time. Unlike traditional governance models that rely on periodic reviews and updates, WDAG can tap into a continuous stream of data from a wide range of stakeholders. This ensures that the ethical frameworks guiding AI development are always in alignment with current societal values.

By analyzing real-time data on community sentiment, for example, of experts who are tasked with the dynamic ethical supervision and guidance of evolving LLM systems, WDAG systems can detect shifts in public opinion regarding what is considered ethical or acceptable behavior for AI systems. Such evolving sentiment, in turn, can provide feedback effects and guidance for AI governance metrics in diverse LLM applications. By staying attuned to these shifts, WDAG ensures that AI governance remains relevant and responsive.

The proposed WDAG system can automate the process of integrating evolving ethical guidelines and legal standards into the AI development lifecycle. This is achieved by maintaining an up-to-date repository of guidelines and standards that are directly applied during the design, development, and deployment stages of AI systems. This real-time updating mechanism ensures that AI technologies do not outpace ethical and legal considerations.

The proposed WDAG framework facilitates the continuous upgrading of AI governance metrics. As ethical standards evolve and new legal requirements emerge, WDAG can adjust governance metrics accordingly. This ensures that AI systems are assessed against the most current benchmarks, maintaining their ethical integrity and legal compliance.

Through a precedent and citation system, WDAG can automate the enforcement of ethical guidelines and legal standards. By codifying these guidelines into enforceable rules within the AI development and deployment

process, WDAG minimizes the risk of ethical breaches or legal violations. This automated enforcement mechanism is crucial for maintaining trust in AI systems, especially as they become more autonomous and integrated into daily life.

By continuously monitoring and adjusting to evolving ethical and legal standards, WDAG provides a preventive approach to AI governance. This contrasts with reactive models that address issues only after they have arisen. Such a proactive stance is essential in preventing harm and ensuring that AI systems contribute positively to society.

Key Components of WDAG-based AI Governance

In the context of AI governance, vertices (nodes) represent distinct governance elements—such as legal precedents, regulatory requirements, ethical guidelines, or governance rules. Directed edges (arcs) illustrate the relationships or citations between these governance elements, establishing a directional flow that signifies logical or legal precedence and dependencies among them.

Assigning weights to the edges in the graph quantifies aspects such as the relevance, authority, impact, or applicability of each governance element in specific contexts. These weights are crucial for prioritizing certain pathways in the decision-making process, ensuring that the most pertinent and influential guidelines are considered in governance decisions.

The acyclic nature of WDAGs ensures there are no loops within the governance framework, facilitating a clear, unambiguous progression from foundational principles to specific governance outcomes. This characteristic is vital for maintaining the integrity and coherence of the AI governance process.

Benefits of WDAG-based AI Governance

As new legal precedents, regulations, or ethical considerations arise, they can be seamlessly integrated into the existing WDAG structure without necessitating a complete overhaul of the governance framework. This capability ensures that AI governance remains current and responsive to the latest developments in AI technology and societal expectations.

The intricate relationships between various governance elements and their applications across different sectors and technologies can be effectively mapped and navigated using WDAGs. This approach enables stakeholders to understand the governance landscape better, facilitating more nuanced and informed governance strategies.

The weighted edges in the WDAG framework aid in highlighting the most relevant, authoritative, or impactful governance elements for particular scenarios. This feature supports stakeholders in making informed decisions by prioritizing governance pathways that align with current needs, legal requirements, and ethical standards.

Uploading AI Models as Posts in the Precedent Credit System

Integrating AI model governance into the WDAG framework can be achieved by treating each AI model or development milestone as a "post" within the precedent credit system. This process involves:

Documenting key attributes of AI models, including design principles, ethical considerations, intended use cases, and compliance with existing governance rules, as vertices within the WDAG.

Establishing directed edges from these AI model posts to relevant legal precedents, ethical guidelines, or regulatory requirements, indicating how each model aligns with or diverges from established governance pathways.

Assigning weights to these connections based on factors like model impact, ethical significance, or regulatory compliance, facilitating a dynamic assessment of the model's governance alignment.

Through the WDAG-based governance framework, stakeholders can dynamically assess, update, and navigate the complex landscape of AI governance, ensuring that AI development remains ethical, compliant, and aligned with societal values. This decentralized approach leverages the inherent advantages of WDAGs to foster a transparent, accountable, and adaptable governance ecosystem for AI.

AI Model Integration

In the context of federal AI learning models, where AI systems must adapt to diverse operational environments and comply with stringent regulations, the WDAG system can dynamically adjust to new ethical guidelines and legal standards. This is crucial in federal contexts where AI applications may range from public safety to health care, requiring a flexible yet robust governance system to ensure that all AI deployments are ethically sound and legally compliant.

By treating each AI model as a post within a WDAG framework, stakeholders can create a comprehensive, visually intuitive map of how well the AI aligns with required governance frameworks. This method allows for ongoing adjustments to governance as AI models evolve and as legal and ethical standards change, ensuring continuous compliance and ethical alignment. This dynamic, structured approach provides a clear, accountable method for managing the complexities of AI governance across various industries and applications. The WDAG system automates the integration of evolving ethical guidelines directly into the AI development lifecycle. This continuous dynamic updating mechanism ensures that AI technologies are always in step with the latest ethical and legal considerations, preventing them from outpacing the regulatory frameworks meant to govern them.

The WDAG governance system is predicated on a decentralized network that captures and integrates community sentiment and ethical

considerations, providing real-time responsiveness that traditional governance models lack. Traditional models often depend on periodic reviews and are slow to adapt, whereas the WDAG system ensures that the ethical frameworks guiding AI development constantly reflect current societal values.

AI governance elements such as legal precedents, regulatory requirements, and ethical guidelines are represented as vertices (nodes) in the WDAG. The directed edges (arcs) between these nodes establish logical or legal precedence and dependencies, which are crucial for understanding the governance structure. Weights assigned to these edges quantify aspects like relevance and impact, prioritizing certain governance pathways over others in decision-making processes.

The acyclic nature of the WDAG ensures that the governance framework is devoid of loops, which facilitates a straightforward progression from foundational principles to specific governance outcomes. This characteristic is vital for maintaining the integrity and clarity of the AI governance process, preventing backtracking and ensuring that each step builds upon the previous one in a logical manner.

As new legal precedents and ethical guidelines emerge, they can be swiftly incorporated into the existing WDAG structure. This capability ensures that AI governance remains contemporary and reactive to the latest developments in AI technology and societal expectations.

The relationships between various governance elements and their application across different sectors and technologies can be effectively mapped using WDAGs. This detailed mapping aids stakeholders in understanding the governance landscape more thoroughly, facilitating more nuanced and informed governance strategies.

The weighted edges within the WDAG framework assist stakeholders in identifying the most relevant, authoritative, or impactful governance elements for particular scenarios. This prioritization helps in making informed decisions that align with current legal requirements and ethical standards.

Importantly, integrating AI model governance within the WDAG framework involves treating each AI model or development milestone as a "post" within the precedent system. Key attributes of AI models, such as design principles and compliance with governance rules, are documented as vertices. Directed edges from these model posts to relevant legal and ethical precedents demonstrate how each model aligns with established governance pathways. Weights assigned to these connections help dynamically assess the model’s compliance and ethical alignment.

CONCLUSION

In the rapidly evolving landscape of AI, the establishment of robust governance frameworks is crucial. As AI technologies grow increasingly complex and become integral across various sectors, the necessity for adaptive governance mechanisms that can evolve alongside these technologies becomes apparent. Traditional governance models, which often depend on static, predefined rules, struggle to keep pace with the rapid development of AI and the intricate challenges it introduces. These conventional approaches, typically reactive and limited to ex-post solutions, are increasingly seen as inadequate for managing the dynamic nature of AI technologies.

The innovative AI governance system proposed in this paper leverages decentralized web3 community governance alongside federated communication platforms, creating a sophisticated and forward-looking framework for the proactive and participatory oversight of AI development. Central to this system is a federated forum platform organized as a WDAG and specialized smart contracts designed for task management and validation. This structure not only promotes real-time consensus building and decision making through web3 community governance but also underpins a scalable and transparent communication network. Within this framework, Validation Pools and Reputation tokens play essential roles in ensuring that the governance system remains continuously updated and responsive, accurately reflecting the collective decisions and ethical standards upheld by the community.

The efficacy of this governance system is exemplified in its application to sectors such as medical diagnosis AI and autonomous driving AI. In these applications, each stage of development is represented as vertices in the WDAG, capturing essential compliance and operational metrics. Directed edges in the graph connect these stages to pertinent legal and ethical standards, with weights assigned to underscore critical areas for compliance and safety. The inherent dynamism of the WDAG facilitates seamless updates and the integration of new regulations or ethical guidelines, thereby ensuring that AI governance remains relevant amidst technological advancements and societal changes. Consequently, this model not only advances AI systems technologically but also ensures they are ethically sound and legally compliant, effectively harmonizing innovation with responsible governance. This comprehensive approach provides a promising outlook for the future of AI governance, suggesting a pathway that could potentially mitigate risks while enhancing the beneficial impacts of AI on society.