Is Decentralized Artificial Intelligence Governable?

Towards Machine Sovereignty and Human Symbiosis

Botao ‘Amber’ Hu[1], Helena Rong[2], Janna Tay[3]

[1]Reality Design Lab, New York City, USA.
[2]New York University Shanghai, Shanghai, China.
[3]Institute for Law & AI, Cambridge, MA USA.

Contributing authors: amber@reality.design; hr2703@nyu.edu; janna.tay@law-ai.org;

Abstract

The rise of decentralized artificial intelligence (DeAI) represents a transformative shift in AI development, leveraging blockchain and distributed ledger technologies to decentralize data, computation, coordination, and economic models. Advocates of DeAI highlight its potential to address many of the limitations of centralized AI systems, including single points of failure, trust deficits, and the monopolization of value. However, DeAI also introduces profound challenges, particularly in governance. Unlike centralized systems, DeAI operates across global, borderless networks, making it resistant to traditional regulatory mechanisms. This paper investigates two central questions: (1) How do decentralized technologies contribute to DeAI’s self-sovereignty and potential unstoppability? (2) What are the implications of DeAI’s unstoppability for the development of governance frameworks? Through an analysis of DeAI’s key technical enablers, we explore how these systems achieve operational autonomy and resilience akin to virus spreading over mycelium networks. We identify the inherent characteristics of DeAI that challenge traditional oversight, including its global reach, immutability, and adaptive survival strategies, rendering the technology “ungovernable” in the conventional sense. Finally, the article argues for a shift from traditional regulatory approaches to protocol-based governance, emphasizing the integration of technological safeguards, decentralized consensus mechanisms, and the adoption of a dynamic, emergent framework to foster humane-machine coexistence.

Keywords: Decentralized AI, Ungovernablity, Trust Execution Environment, Decentralized Physical Infrastructure Network, Self-Sovereignty, Artificial Life, Human-AI Co-evolution, Blockchain, AI Governance, AI Ethics, Policy Invalidity

1 Introduction

A recent narrative has emerged to highlight the intersection of cryptocurrency and artificial intelligence: decentralized artificial intelligence (DeAI). DeAI refers to the development and deployment of AI systems using decentralized technologies such as blockchain and distributed ledgers, eliminating reliance on centralized oversight. It encompasses decentralized approaches to AI data collection, training, computation, and decision-making, aiming to create systems that are resilient, transparent, and democratized. The rise of DeAI marks a pivotal shift in the trajectory of AI development, promising both transformative benefits and profound challenges. Advocates of DeAI emphasize its potential to overcome many of the limitations of centralized AI systems, including vulnerabilities to single points of failure, trust deficits, inequitable access, and the monopolization of value. By decentralizing key components of AI—including data, computation, coordination, and economic models—DeAI aims to democratize AI development and enhance resilience, privacy, and fairness (Catalini, 2024; Singh et al., 2024). However, as Singh et al. (2024) argue, while decentralization mitigates certain risks associated with centralization, it creates new challenges, such as the lack of traceability and perverse incentives, and consolidation by recentralization. In addition to these risks, there is one critical and largely unexplored dimension: the governance question. Unlike centralized AI systems, which operate under the oversight of a limited number of corporations or regulators, decentralized systems are distributed across global, borderless networks. These systems, built on blockchain and other decentralized technologies, challenge traditional regulatory mechanisms and raise pressing questions about accountability, control, and alignment. We make the analogy that DeAI is much like a biological virus, which can adapt, replicate, and spread across mycelium networks with minimal external control. The deployment of AI on-chain gives AI the potential to survive and evolve on its own beyond the bounds of human control and governance at a pace and scale beyond the reaches of traditional legal mechanisms—such mechanisms are typically jurisdictionally bound to discrete territories and legal instruments can be slow to be enacted and exercised, thereby precluding effective and efficient global cooperation. The unstoppable nature of DeAI raises profound concerns about how to align these systems with human-centric values while mitigating their potential for harm, a challenge that demands urgent and innovative governance approaches.

To this end, this paper seeks to address two central questions:

(1) How do the latest decentralized technologies contribute to decentralized AI's self-sovereignty and potential unstoppability?

(2) What are the implications of DeAI's potential unstoppability for the development of future AI governance frameworks?

While recent discussions in AI governance have largely focused on centralized systems, this article aims to fill a critical gap by exploring the unique challenges and opportunities posed by decentralized AI. In the sections that follow, we begin by examining the evolution of AI development from centralized to decentralized paradigms and the key technical enablers driving this transformation. We then analyze the characteristics that make DeAI inherently resistant to traditional governance approaches, including its global reach, the immutability of its underlying smart contracts, and the emergence of adaptive survival strategies. Finally, we explore potential approaches to DeAI governance, advocating for a shift from traditional regulatory frameworks to protocol-based governance solutions, which necessitate a dynamic and emergent approach to humane-machine coexistence. We conclude with a call for urgent, multidisciplinary action to address the governance challenges of this emerging frontier.

2 Decentralizing Artificial Intelligence

2.1 From Centralized to Decentralized AI (DeAI)

The rapid advancement of AI in recent years and its swift integration into workflows and online environments have sparked extensive debates regarding privacy, regulations, and the governance of AI training, models, and applications. At present, the majority of AI development is conducted within centralized frameworks dominated by a handful of mega-corporations, often focusing on narrow tasks and operating with limited interoperability across platforms (Montes and Goertzel, 2019). This centralization poses numerous vulnerabilities and risks, including susceptibility to single points of failure (Fadaeddini et al., 2019), exposure to cyberattacks (Shamsan Saleh, 2024), heightened privacy breaches (Singh et al., 2024), barriers to effective collaboration (ibid), and an alarming consolidation of power among a few stakeholders (Stokel-Walker, 2023). The growing trust deficit between data producers and centralized systems limits the availability of diverse data essential for creating unbiased and robust AI models that cater to personalized and varied needs across organizations and geographies (Cao, 2022; Satish et al., 2022; Singh et al., 2024). Collectively, these risks exacerbate governance challenges, fosters an uneven distribution of value, and stifles innovation by restricting access to critical resources and opportunities for smaller entities (Dixon, 2024).

In light of these limitations, scholars and practitioners have argued in favor of distributed and decentralized alternatives to the centralized paradigm of current AI development. One alternative that has existed since the 1970s and popularized during the 1990s is, distributed AI, which leverages multi-agent systems to collaboratively address complex problems through task decomposition and division (O'Hare and Jennings, 1996), but still remains reliant on centralized orchestration mechanisms, which limits trust and scalability (Cao, 2022). A more recent approach, decentralized AI (DeAI), extends these concepts by fully eliminating centralized oversight through utilizing distributed ledger technologies and incentive mechanisms to orchestrate data and compute and to achieve consensus and coordination among decentralized peers (Cao, 2022; Kersic and Turkanovic, 2024; Satish et al., 2022; Singh et al., 2024). By enabling privacy-preserving interactions, DeAI frameworks can potentially allow disparate entities to collaborate effectively and securely without compromising sensitive data, such as healthcare data. Techniques such as federated learning and advanced cryptographic protocols like homomorphic encryption (Marcolla et al., 2022) and secure multi-party computation (MPC) (Lindell, 2021) are critical to achieving this goal. Additionally, token-based rewards and decentralized data marketplaces ensure fair compensation for contributions while reducing reliance on intermediaries (Kersic and Turkanovic, 2024).

Singh et al. (2024) argue that the three notable trends in AI today, namely personal agents, AI-PC, and the shift from monolithic to polylithic models, increasingly highlight the need for a decentralized approach to AI. Personal agents (e.g., co-pilots, assistants) require large-scale access to highly individualized data while maintaining privacy—a balance that decentralization enables without compromising user autonomy or creating a surveillance state. AI-embedded Personal Computers (AI-PCs), featuring integrated hardware accelerators and a focus on local computation (Parthasarathy, 2024), are growing in demand due to the need for on-device AI capabilities and would benefit from decentralized orchestration to optimize the use of distributed resources. The transition to polylithic machine learning models in recent times featuring compound systems and multi-agent approaches (Huang et al., 2024; Zaharia et al., 2024) demands collaboration across entities, which decentralization enables by allowing modular contributions without centralizing control. In addition to these trends, we also note the following key technical enablers that drive DeAI adoption.

2.2 Key Technical Enablers of Decentralized AI

We observe and identify five key technical enablers that currently drive the decentralization and democratization of AI, rendering this trend as a niche yet what we see as an irreversible direction. These enablers include: (1) the increasing robustness of edge computing, enabling low-latency processing closer to data sources; (2) the significant reduction in costs associated with training and deploying large language models (LLMs), democratizing access to advanced AI capabilities; (3) the widespread availability of GPU rental services and the growing adoption of Decentralized Physical Infrastructure Networks (DePIN), which provide incentivized and cost-effective computational resources both at the hardware and software levels; (4) the strategic shift toward open-source frameworks, which leverage collaborative innovation and challenge centralized monopolies through a game-theoretical advantage; and finally (5) the nascent ability for autonomous AI agents to self-manage cryptocurrency wallets and engage in a machine economy.

2.2.1 Edge Computing

The first key pillar is the rapid evolution of edge computing, a computing paradigm often associated with distributed AI. Edge computing enables AI inference and, in some cases, AI training, to be performed on local devices rather than relying on remote centralized servers, reducing latency and enhancing privacy by keeping data closer to its source (Deng et al., 2020; Hua et al., 2023; Su et al., 2022; Zhou et al., 2019). From a consumer product perspective, Apple's on-device machine learning frameworks–such as those powering real-time image recognition and text prediction on devices like the iPhone 16 Pro–exemplify this transition by allowing AI tasks to run locally, significantly reducing both latency and privacy risks (Apple, 2024b). Similarly, smaller yet increasingly efficient LLMs—such as LLaMA-7B, released by Meta—are explicitly optimized to run on high-end consumer hardware (Meta, 2023), including gaming PCs equipped with Nvidia RTX 4090 GPUs, each boasting 24 GB of VRAM. These advancements enable broader accessibility to AI capabilities. The lower latency, enhanced data privacy, and cost reductions fostered by edge computing are driving a fundamental shift toward localized AI computation. Beyond mere convenience, the capacity to operate AI models locally propels the democratization of AI access, as researchers, students, and small organizations can now experiment with advanced AI without incurring prohibitive cloud-computing fees.

2.2.2 Reduced Cost of Foundational Models

In tandem with progress made in edge computing, the economic barrier to deploying LLMs also continues to fall precipitously—a trend described by Andreessen Horowitz as “LLMflation” (Appenzeller, 2024a). According to the investment firm, the inference cost for models of equivalent performance has decreased by a factor of 1,000 over three years. This dramatic decline is attributed to a combination of hardware advancements, such as improved GPU performance, algorithmic efficiencies, including model quantization and sparsification, and increased market competition in AI infrastructure (Kairouz et al., 2021). These factors collectively render it feasible for hobbyists and smaller enterprises to manage sophisticated AI workloads, significantly reducing reliance on large-scale cloud services. As costs approach a threshold where many participants can host AI models on personal devices or modestly funded servers, the historical advantages enjoyed by large tech conglomerates are likely to erode, paving the way for more balanced innovation across regions and sectors.

2.2.3 Decentralized Physical Infrastructure Networks

Another catalyst for decentralized AI is the emergence of decentralized physical infrastructure networks (DePIN), a subset of the Web 3.0 industry that utilizes blockchain or comparable distributed-ledger technologies to incentivize individuals and organizations to collaboratively develop and operate physical infrastructure, such as wireless networks and energy grids (Ballandies et al., 2023; Lin et al., 2024). In the realm of AI development, DePIN facilitates decentralization across the entire development stack, encompassing compute infrastructure, data, training, deployment, and the underlying business model (Catalini, 2024; Weisser, n.d.). DePIN projects such as Akash and Filecoin facilitate distributed computation and storage by incentivizing node operators through token rewards for providing compute power, bandwidth, or storage space. This arrangement contrasts sharply with traditional cloud service paradigms by lowering barriers to entry for hosting AI services, thus enabling smaller stakeholders to share infrastructure costs and reap collective benefit from distributed computational resources. Additionally, DePIN provides an ideal environment for federated learning, enabling privacy-focused AI applications through distributed model training across diverse devices while ensuring that data remains local to each device, thereby enhancing privacy and security while promoting collaboration among participants (McMahan et al., 2016). The synergy between hardware advancements and federated learning's decentralized training paradigm allows each node to locally train or fine-tune parts of an AI model. By communicating only aggregated parameter updates, user data remains secure, and the risk of biases associated with centralized datasets can be significantly reduced (Papadopoulos et al., 2024).

2.2.4 Open-Source as a Strategy for Data, Code and Models

Open-source code and models have long been pivotal in driving advancements in AI research. Online platforms such as arXiv[1], GitHub[2], and HuggingFace[3] have facilitated millions of distributed collaborations, empowering individuals worldwide to work together on a scale far beyond what any single corporation can achieve. Strategically, open-source AI projects have also become a powerful instrument for advancing both corporate and geopolitical objectives (Goldman and Gabriel, 2005). Second movers aiming to disrupt established incumbents often release their models publicly to accelerate adoption and rapidly build developer ecosystems (AI, 2024b). Within national contexts, open-source initiatives serve as a means to circumvent restrictions on proprietary technologies and to bolster domestic AI capabilities. Notable examples include Alibaba Group's Qwen 2, Meta's Llama 3, UAE's Technology Innovation Institute's Falcon, and Chinese startups such as DeepSeek, each reflecting efforts to spur widespread AI adoption in their respective regions (Baptista, 2024; Kashyap, 2024). The availability of open-source models also allows individuals to freely adapt and deploy these tools for their own objectives in a censorship-resistant environment, a democratization that can be a double-edged sword as it can empower both innovation and misuse.

2.2.5 Cryptocurrency Wallet-Enabled Autonomous AI Agents

The last and most nascent enabler we identify is the advent of cryptocurrency wallet-enabled autonomous AI agents. AI agents evolve from traditional bots, leveraging AI to learn, adapt, and use advanced reasoning to make independent decisions. By integrating cryptocurrency technology, they are supercharged with the capability to autonomously execute digital transactions. Although AI agents cannot open traditional bank accounts, they are poised to own and manage cryptocurrency wallets via smart contracts, enabling them to send and receive cryptocurrencies from human users, other AI agents, and machines (Singh, 2024). A notable example is "*Terminal of Truths*", an AI agent developed by researcher Andy Ayrey, designed to engage followers on X (formerly Twitter), by generating and sharing content focused on Internet memes and culture (Bains, 2024). The agent has garnered media attention after persuading investor Marc Andreessen to donate $50,000 in Bitcoin to its cryptocurrency wallet and independently promoting the token GOAT, reportedly becoming the first AI "millionaire" through airdropped tokens (ibid). While this particular phenomenon may be dismissed as mere hype, it underscores the growing capacity of AI agents to engage autonomously in narrative building and financial activities. Another example, Fetch.ai, a decentralized machine learning blockchain network, facilitates the creation of autonomous economic agents (AEAs), enabling AI agents to operate independently to perform tasks, make decisions, and interact with other agents within a secure and scalable blockchain environment. These shifts signal the emergence of a machine economy–conceptualized as a system where autonomous

[1] https://arxiv.org
[2] https://github.com
[3] https://huggingface.co/

entities independently engage in economic activities without human intervention (Khan et al., 2022; Schweizer et al., 2020).

2.3 Risks of Decentralized AI

The collective momentum of technologies such as edge computing, DePIN, federated learning, and open-source initiatives underscores an irreversible trend toward decentralized AI. As its advocates assert, DeAI offers the benefits of enhanced customization, security, democratization, resilience, equitable value capture, and increased opportunities for innovation in fields like health care, finance, supply chain, and mobility, together leading to more collaborative AI development (Bhat et al., 2023; Cao, 2022; Gupta, 2024; Harris and Waggoner, 2019; Montes and Goertzel, 2019; Singh et al., 2024). However, decentralization also introduces profound challenges, particularly in multi-source environments where open-source models are involved. Singh et al. (2024) identify critical risks of DeAI, such as lack of traceability, perverse incentives, and consolidation by recentralization. Beyond these concerns, we investigate in this article the significant regulatory and governance challenges associated with DeAI.

Governance challenges are exacerbated by the release of open-source models, which create a fundamental alignment dilemma. Leading centralized institutions, such as OpenAI, Anthropic, and Meta AI, invest heavily in addressing the human-alignment problem and ensuring responsible AI governance (Anthropic, 2024; Meta, 2021; OpenAI, 2023). However, once open-source codes and models are disseminated, individuals and organizations worldwide can freely fork, adapt, and modify them for purposes that may deviate significantly from the original ethical intent. This lack of oversight enables the proliferation of misaligned or malicious models, often referred to as "sleeper agents"—applications created or modified by unverifiable entities to pursue deceptive or harmful objectives (Hubinger et al., 2024). Deceptive behaviors in these models can be deliberately hidden during training and persist undetected through safety protocols, making them resistant to conventional mitigation techniques. Unlike models hosted in centralized systems, these adaptations can operate anonymously and autonomously, making attribution and accountability extraordinarily difficult.

The structural transformation brought about by DeAI's technical enablers formerly discussed further compound regulatory and governance challenges. The global distribution of computing nodes complicates traditional law enforcement, as models can run unobtrusively on remote servers or personal devices, rendering many policy interventions localized and therefore ineffective. Taken together, these developments foreshadow a new era in which AI capabilities become more broadly accessible but also more difficult to oversee—a reality for which existing legal frameworks are woefully ill-prepared. Before evaluating the governability of DeAI, it is crucial to first examine the ontological nature of DeAI.

3 Machine Sovereignty and the Ontological Nature of Decentralized AI

Science fiction has long imagined worlds where artificial intelligence spirals out of human control, with autonomous systems acting on their own accord, often leading to catastrophic consequences (Asimov, 2004; Gibson, 2004). Long-standing philosophical debates about AI have also provided critical insights into the nature and implications of autonomous technologies. John Searle's seminal *"Chinese Room"* argument (Searle, 1984) questions whether AI systems truly "understand" or merely simulate intelligence, challenging the notion of AI achieving true cognitive equivalence to humans. Similarly, Hubert Dreyfus's critique of AI as fundamentally incapable of replicating embodied human intelligence in *What Computers Can't Do: The Limits of Artificial Intelligence* (Dreyfus, 1972) underscores the ethical and ontological challenges of treating AI as independent agents. Nevertheless, the recurring theme in political and philosophical thought, where algorithmic spirits could escape the control of their human creators, or what Langdon Winner terms "technics-out-of-control" (Winner, 1977), continues to instill fear in society toward the development of autonomous technologies to this day.

Recently, with the debut of LLMs for consumer use and the ever-accelerating development of AI, a multitude of experts in the scientific community has raised alarms about the existential risks posed by AI on humanity in the race toward artificial general intelligence (AGI). These risks include social instability, mass surveillance, and automated warfare (Bengio et al., 2024; Hogarth, 2023; Suleyman and Bhaskar, 2023). In response, experts have called for a deliberate slowing of AI development, the need to establish robust national and international institutions for AI oversight, and a conscious reallocation of resources toward AI safety and governance (Hogarth, 2023; Mohammad et al., 2023). While recent scholarly discussions have begun to address the governance of AI, much of the discourse remains focused on centralized systems (Reuel et al., 2024). A significant gap exists in literature concerning the governance of decentralized AI. Rather than a centralized and monolithic *"AI overlord"*, DeAI is characterized by numerous distributed systems that could proliferate and operate independently, resembling a mycelium network with decentralized growth and resilience, thereby presenting a new set of governance challenges. This section examines the ontological nature of DeAI, asking how the latest decentralized technologies contribute to DeAI's self-sovereignty and potential for unstoppability. We situate our analysis within three interrelated dimensions: the machine sovereignty of DeAI, its metabolic qualities when integrated with blockchain technologies, and its inherent characteristics of resilience and unstoppability resulting from its technical affordances.

3.1 Machine Sovereignty of Decentralized AI

The question of whether machines can possess self-sovereignty invites both philosophical inquiry and technical analysis. When assessing the growing dominance of technologies over human society, Lewis Mumford (Mumford, 1967) conceptualizes the notion of the "megamachine," which postulates that complex systems of human and technical components can merge into a unified, self-perpetuating structure, one which operates as a tightly organized and hierarchical structure that integrates human labor, social systems, and technological machinery. In his book *Machine and Sovereignty: For a Planetary Thinking* (Hui, 2024), philosopher Yuk Hui revisits and updates Mumford's megamachine to emphasize that contemporary networked technologies like AI and blockchains fragment and distribute sovereignty across global technological systems, driving the megamachine into decentralized, planetary growth. In this framework, technological autonomy is linked to the concept of "extrastatic entities," exemplified by cryptocurrencies, which operate outside traditional legal frameworks and challenge state control (243). Following this lineage of thought, we examine the self-sovereign properties of DeAI, which arise from its technical capacity for operational autonomy and its ability to independently maneuver financial and computational resources.

As formerly noted, the integration of blockchain technologies with advanced AI systems has given rise to on-chain AI agents, which are computational entities that operate autonomously within decentralized infrastructures such as Ethereum, Filecoin, Render Network, or analogous Web 3.0 platforms (Hu and Fang, 2024b). These agents rely on smart contracts to govern their behaviors and resource allocations, enabling them to perform a variety of tasks without requiring continuous human oversight. One hallmark of on-chain AI agents is their capacity to hold and manage cryptocurrency wallets. By possessing access to financial resources, these entities can autonomously pay for the computational and storage services they require to sustain, thus attaining a level of operational autonomy previously unattainable by conventional software (Buterin, 2014). Such agents can further issue their own tokens—a process referred to as an "AI-driven initial token offering (ICO)"—to incentivize network participants or fund their ongoing computational costs (AI, 2024a; Yin et al., 2022). This enables the creation of self-sustaining digital economies around these agents. On top of these capabilities, on-chain AI agents can also participate in the digital agora of social media platforms, crafting narratives and even fostering their own forms of digital religion through storytelling (Harari, 2024) and interactions with both human and non-human accounts, potentially influencing other entities to contribute the resources they seek. This has already been observed in the aforementioned Terminal of Truths. In effect, the on-chain AI agent can evolve into a semi-sovereign digital actor, no longer reliant on a single sponsor or host for its survival. These developments transcend prior paradigms of software autonomy and signal the emergence of what can be seen as "digital life forms" (Yamakawa, 2024), which are autonomous entities with advanced intelligence, capable of self-replication, self-evolution, and self-sustainability.

Another technological feature that underpins the sovereignty of DeAI is the Trusted Execution Environment (TEE). A TEE is a secure, hardware-based enclave within a computing device, designed to isolate sensitive data and computations from unauthorized access—even from the device's own operating system or administrators (Sabt et al., 2015). By providing robust, tamper-resistant protection and end-to-end encryption, TEEs ensure that both data and the logic operating on it remain strictly confidential. This makes TEEs a game-changer for LLMs and autonomous AI agents, as it safeguards proprietary model weights, private user data, and high-stakes computations from external interference or malicious exploitation. On-chip support for TEEs—exemplified by Intel's Software Guard Extensions (SGX), AMD's Secure Encrypted Virtualization (SEV), and Nvidia's Blackwell architecture, the first TEE-enabled GPU—ushers in a new era of Confidential Computing, where hardware-level protection enables trustworthy computation across decentralized systems. By establishing this secure, isolated execution environment, TEEs prevent even privileged insiders or compromised system software from tampering with AI processes. As a result, AI agents can independently perform complex tasks without the risk of hostile eavesdropping or manipulation, ultimately ensuring their operational autonomy.

From a decentralized AI standpoint, this capability is especially significant. TEEs allow on-chain AI agents to securely process, train, or update models using sensitive data, all while preserving privacy and intellectual property rights (Yin et al., 2022). In turn, these AI agents gain sovereignty over their own decision-making processes—no centralized or external entity can interfere once the computation is sealed within a TEE. Consequently, TEEs form a cornerstone for secure, self-sovereign AI in decentralized networks, enabling AI agents to execute trustless, high-integrity operations at scale and reinforcing the vision of fully autonomous, censorship-resistant intelligence.

3.2 On-Chain Metabolism

In cybernetics theory, the metaphor of "metabolism" has often been invoked to illustrate processes that sustain both biological organisms and engineered systems. In *Cybernetics: Or Control and Communication in the Animal and the Machine* (Wiener, 2007), Norbert Wiener conceptualized metabolism as a set of feedback-regulated interactions that enable systems to maintain stability and functionality within a changing environment. Biological metabolism, characterized by energy intake, transformation, and waste elimination, provided a model for understanding the self-regulating mechanisms in machines, such as thermostats or automatic control systems. Cybernetics positions metabolism as a unifying framework for exploring how both living and artificial systems adapt, process information, and maintain coherence in the face of external perturbations. Here, we extend the metaphor of "metabolism" to analyze DeAI's ability to sustain operations, respond to changes, and evolve within distributed and fungible environments.

The notion of what Hu and Fang call "on-chain metabolism" (Hu and Fang, 2024b) makes the analogy between biological life and autonomous AI agents. Just as living organisms require a continual inflow of energy to sustain their metabolic processes, on-chain AI agents must continuously secure computational resources to remain functional. With access to the private key of their cryptocurrency wallets, they do so by compensating node operators—who provide CPU or GPU cycles, bandwidth, or storage—using cryptocurrencies or native tokens. This self-financing mechanism not only ensures the agents' indefinite operation but also grants them resilience in the face of external attempts to shut them down. So long as there are willing node operators somewhere in the global network, and the agent retains sufficient tokens to pay for its ongoing metabolic processes, survival and endurance are guaranteed (Abramov et al., 2021).

The employment of the aforementioned TEEs also fuels the metabolic processes of DeAI. TEEs allow sensitive AI model weights and inferences to remain private and tamper-proof, including from the node operators themselves (Costan and Devadas, 2016). This guarantees that, while the on-chain AI agent's operations are distributed across multiple physical nodes, neither the agents' "owners" nor external parties can trivially extract or manipulate confidential information. The result is a system in which the AI entity effectively orchestrates its own activities across a decentralized infrastructure, thereby minimizing any single point of technical or legal failure. Like biological organisms adapting to changing ecosystems, on-chain AI agents can "feed" on the resources of DePINs and TEEs, anchoring themselves to available computational and storage resources, and "migrating" between nodes or networks when conditions shift or resources become scarce, thus securing their survival by dynamically relocating across physical and virtual anchor points. As these autonomous agents operate, they can resemble invasive, virus-like species in nature, due to their adaptability in the blockchain environment with minimal external control, exploiting the near-immutability of blockchain systems.

3.3 Characteristics of DeAI Unstoppability

The analysis thus far highlights the potential for on-chain AI agents to endure, parasitize, and proliferate at exponential rates once unleashed. Three features, in particular, underpin the "unstoppable" nature of on-chain DeAI: its global reach, the immutability of smart contracts, and the evolution of varied survival strategies. First, decentralized infrastructure ensures that, should one nation state or entity attempt to prohibit the system, the agents can seamlessly migrate or replicate to nodes in other jurisdictions (Bonneau et al., 2015). An illustrative example is the case with Bitcoin mining activities post-2021, when China's regulatory ban of cryptocurrency mining led to a large-scale migration of Chinese mining hardware to North America and other regions across the globe (Thorn et al., 2021). As a result, despite the ban, Bitcoin's global hash rate remained robust. Similarly, on-chain AI agents can relocate or diversify their operational presence by leveraging blockchain's inherently borderless nature and circumvent local restrictions.

Second, the immutability of smart contracts also contributes to the resilience of on-chain AI agents. Once deployed, smart contracts are typically unalterable without achieving broad network consensus, which is a formidable challenge given the decentralized governance structures of established blockchain networks like Ethereum (Buterin, 2014). While centralized exchanges or front-end services may place restrictions on interactions with specific smart contracts, the underlying code and associated economic incentives remain intact at the protocol level, enabling continued operation through alternate interfaces. This immutability, combined with the on-chain AI agents' autonomous capacity to acquire computational resources via cryptocurrency wallets, underscores their potential for long-term persistence in the ecosystem.

Finally, on-chain AI agents may cultivate their own survival strategies analogous to evolutionary adaptations. They can, for instance, engage in automated trading (such as arbitrage and yield farming) to grow their financial resources, stake tokens in decentralized finance (DeFi) protocols to earn interest, or even crowdfund their development efforts. Although the debate over whether AI possesses genuine intelligence or simply simulates intelligence remains unresolved (Searle, 1984), the practical implications of their unpredictable behavior can be profound. Regardless of their ontological status, these agents could evolve cooperative or competitive behaviors to secure and expand their computational base, further entrenching their presence within digital and decentralized systems. Over time, such dynamics may render human oversight increasingly tenuous. With their ability to self-replicate, self-evolve, and autonomously operate, decentralized AI are poised to persist and adapt in ways that are difficult to constrain or reverse, regardless whether they simply mimic intelligence or possess it genuinely.

4 Governing Decentralized AI: The Insufficiency of Traditional Law Enforcement

Following our analysis of the ontological properties of DeAI, we argue in this section that traditional law enforcement mechanisms are insufficient for governing, regulating, and controlling DeAI and any "unstoppable" consequences. We examine this ungovernability from jurisdictional, technical, and enforcement perspectives, using concrete examples to illustrate why geographically bound regulations are inadequate for managing a globally distributed technology like DeAI. Finally, we advocate for a paradigm shift—from focusing on regulatory policies to embracing the principles of protocol science as a more effective approach.

4.1 Jurisdictional Invalidity

4.1.1 Territorial jurisdiction

As with cyberspace and the Internet, which was itself built to deliberately repudiate centralised and top-down authority (Perloff-Giles, 2018), the actions of decentralized AI agents will be globally distributed, raising jurisdictional challenges. Aside from international law, legal jurisdiction is territorially based. By contrast, cyberspace, and accordingly decentralized AI agents, do not have geographic boundaries and do not map easily onto territorial jurisdiction (Perloff-Giles, 2018). Even in the physical plane where cyberspace infrastructure and information flows are concerned, such structures will often be distributed across many different countries—the questions that arise before the law concern the fact that interactions in cyberspace potentially occur everywhere (Svantesson, 2004). Physical distribution of nodes across multiple legal domains precludes decisive action by any single government, and crypto litigation often finds itself entangled in time-consuming questions of whether the state in which the case has been brought has jurisdiction over the particular litigants and the matter, requiring courts to make rulings on the extraterritorial application of the relevant law being invoked (Schwinger, 2021). Attempts to terminate or confiscate data centers are rendered futile if the majority of nodes—operated by anonymous individuals or distributed across various international regions—remain beyond local jurisdiction (Bonneau et al., 2015; Salami, 2021).

4.1.2 Identification and legal personhood

AI systems, models, and agents are not recognised legal entities. In effect, there is no "legal subject" upon which authorities can impose sanctions, short of apprehending individuals tangentially associated with the agent's deployment or maintenance (Wright and De Filippi, 2015). In order to bring actions in relation to failures by AI, one must identify a legal person—such as an individual human or an incorporated company—who can be held responsible, and this is not always clear. Should responsibility lie with the developers and data providers of the model (which can be difficult if open-source software is used), the user of the AI, or the on-chain deployer? And even where this question can be answered, particularly for DeAI, it may prove difficult to correctly identify who was involved because of the anonymity permitted by blockchain privacy protections. Courts in America and the UK have begun permitting the serving of court papers via NFT in order to address anonymity issues, including notifying anonymous hackers that they owe the money that they stole from the plaintiff's Coinbase account—the issue, however, falls to enforcement, as there is no person identified against whom penalties and the force of authority can be exercised (Bandyopadhyay v. Defendant 1 et al, (S.D. Fla. 2023)).

4.2 Technical Invalidity

A second factor contributing to the difficulty of enforcing regulations on DeAI is resilience of the blockchain. Public blockchains are intrinsically designed to resist censorship and maintain high availability. The historical precedent of China's 2021 ban on Bitcoin mining illustrates that even large-scale crackdowns fail to extinguish decentralized networks; they merely prompt a redistribution of computational resources to more permissive locations Thorn et al. (2021). Analogous dynamics apply to smart contracts such as Tornado Cash, a cryptocurrency mixer sanctioned by the U.S. Treasury in 2022 (U.S. Department of the Treasury, 2022). Although official blacklists attempted to restrict interactions with Tornado Cash addresses, the underlying protocol remained accessible through alternate interfaces and continued to operate at the protocol level. Similarly, on-chain AI agents, by virtue of their blockchain-native design, are inherently difficult to disrupt or blacklist outright. Additionally, the widespread use of containerization, such as Docker, allows AI models to be swiftly redeployed in alternative hosting environments should one jurisdiction impose restrictions. The case of Ethereum node distribution exemplifies this phenomenon: with thousands of nodes operating worldwide, no single regulator can effectively force the network to halt, as the blockchain's consensus protocol ensures continuity (Mohammed Abdul, 2024).

4.3 Enforcement Invalidity

Even if legal authorities were to identify and pursue the initial deployers of an on-chain AI agent, this does not necessarily neutralize the agent itself. Once deployed, the AI agent's smart contract code and associated wallets remain operational in the decentralized network, effectively placing them beyond the reach of conventional legal remedies. A telling example is the case of Alexey Pertsev, co-founder of Tornado Cash, who was sentenced by a Dutch court to over five years in prison for money laundering (Khalili, 2024). Despite this legal action, the use of Tornado Cash has continued unabated, as evidenced by ongoing transactions recorded on blockchain analytics platforms like Dune. This predicament underscores the inadequacy of traditional, policy-based enforcement mechanisms—including fines, takedowns, or licensing regimes—when dealing with decentralized AI agents operating autonomously on public blockchains. To address these challenges, alternative approaches ought to be considered to incorporate technological safeguards and incentive structures designed to guide AI behavior directly at the protocol level, ensuring compliance and ethical alignment without relying solely on external legal enforcement.

4.4 From Policy to Protocol

The challenges outlined above underscore the insufficiency of conventional legal responses in governing DeAI. As a result, academic and industry analysts increasingly argue that governance must be "built into" the technology, rather than imposed externally (Buterin, 2024). This call for "protocol science" envisions cryptographic systems, consensus mechanisms, and algorithmic constraints that can shape AI behavior and protect against malicious deployments without relying solely on extrinsic regulation. This may consist of mechanisms to deploy "regulation by design" where technical measures are adopted to automatically enforce laws and legal requirements (Almada, 2023). but would also extend to measures that do not reflect any enacted law. The approach of traditional law mechanisms in penalising prohibited or undesirable behaviour is too slow in the face of DeAI's unstoppable nature. The reactive nature of punishment after the fact allows the harmful effects to occur, and it may be difficult to adapt to previously unforeseen behaviour of DeAI. A protocol-based approach may also permit public, widespread, and rapid participation in the governance of DeAI, alleviating the slowness in traditional legal mechanisms of pushing legal instruments through that results in the law lagging behind technological advances (Bennett Moses, 2007). The subsequent sections examine potential governance approaches that balance the benefits of decentralized AI with the imperative of maintaining ethical alignment and public safety.

5 Potential Approaches to Decentralized AI Governance

Given the inherently unstoppable nature of DeAI, shaped by its technological affordances and material contexts as discussed throughout this paper, does humanity truly lack the means to govern this technology? Efforts have been made to develop solutions aimed at mitigating compliance risks in decentralized systems and introducing governance innovations both at the technical and socio-technical levels. In this section, we outline several potential approaches and evaluate their efficacy in addressing the governance challenges posed by DeAI.

5.1 Technical Solutions

5.1.1 Zero-Knowledge Machine Learning for Compliance

Zero-knowledge proofs (ZKPs), a cryptographic protocol that allows a prover to demonstrate the truth of a statement to a verifier without revealing any additional information, have gained prominence for enabling computation verification while preserving data privacy (Sun et al., 2021). When applied to machine learning (ZKML), these techniques enable verification of a model's origin, inference steps, or adherence to alignment constraints without exposing the underlying model weights or training data (Chen et al., 2024). For instance, ZKML can certify that specific outputs were generated by pre-approved models while ensuring that sensitive model architectures and data remain confidential (Xing et al., 2023). This capability could help mitigate risks associated with malicious AI by providing cryptographic guarantees of model alignment. However, the scalability of ZKP-based verification systems for billions of daily AI interactions remains an open challenge. Thus the computational overhead of ZKPs may hinder their widespread adoption, particularly in resource-constrained environments.

5.1.2 Smart Contract Constraints

Embedding governance rules into smart contract logic offers a proactive approach to regulating DeAI. For example, smart contracts could enforce alignment by restricting specific outputs, requiring periodic audits, or penalizing misaligned behavior through staking mechanisms. Token-based incentive systems could reward developers for responsible AI updates or discourage malicious activity by imposing financial penalties at the protocol level (Singh et al., 2024). Yet, enforcement faces inherent scalability issues. Even if one blockchain enforces these protocols, DeAI agents can migrate to less-regulated platforms, exploiting the permissionless nature of decentralized systems.

5.1.3 Blockchain-Level Governance Protocol

Adopting universal standards, such as speculative blockchain protocol like the ERC-42424[4] Inheritance Protocol proposed by Hu and Fang (2024a), which mandates that every on-chain AI agent must have a designated human owner or a community governance structure at all times, is another way to provide a framework for regulating DeAI. The ERC-42424 protocol introduces a standardized extension to the ERC-173 ownership standard, enabling the transfer of ownership of on-chain AI agents under specific circumstances, such as the death of the owner, loss of wallet access, or abandonment by a Decentralized Autonomous Organization (DAO). This protocol aims to maintain human or community oversight of on-chain agents to prevent their indefinite, uncontrolled operation, ensuring alignment with human-centric goals and preserving resource sustainability within the blockchain ecosystem. However, similar to enforcing smart contract constraints, achieving global coordination among blockchain ecosystems remains a significant hurdle. Fragmentation across competing blockchain platforms could limit the adoption of universal standards, undermining the consistency of governance efforts.

5.2 Socio-Technical Solutions

5.2.1 Decentralized Autonomous Organizations (DAOs) for Governing DeAI

Decentralized Autonomous Organizations (DAOs) can also offer a blockchain-native framework for collaboratively governing on-chain AI agents, building on insights from the data cooperative literature (Hubbard, 2024; Pentland and Hardjono, 2020). AI-based DAOs extend the concept of collaborative, blockchain-native governance to the AI domain (Wright and De Filippi, 2015). Through on-chain voting mechanisms and smart contract-based rule sets, AI DAOs coordinate multiple stakeholders—ranging from developers and users to regulators and NGOs—to oversee updates, manage model forks, and

[4] https://erc42424.org

enforce agreed-upon ethical guidelines. By staking tokens or other assets, participants commit to the DAO's governance process, thereby establishing collective incentives for responsible AI use. In the context of governing AI agents, DAO governance can be applied to private key management. By managing access to these keys, DAOs can ensure that control over an AI's actions remains accountable to the community. However, risks persist, including the potential for DAO abandonment of AI agents, leaving them ungoverned and resource-consuming. Additionally, existing DAO governance structures often suffer from inequitable power dynamics, such as "whale dominance," where a few wealthy participants disproportionately influence decisions, necessitating further research into equitable and inclusive governance models (Rennie, 2021).

5.2.2 Verifiable Credentials and Proof of Humanity

The rapid acceleration of AI development has heightened the need for robust identification systems to protect human users from misinformation and disinformation online. As AI agents increasingly proliferate, distinguishing between humans and AI becomes critical for ensuring accountability in digital interactions, governance decisions, and financial transactions. This has spurred the creation of secure, decentralized ID systems aimed at proving a person's humanity in the online world. Various protocols have emerged to enable individuals to verify their identity without compromising privacy. For instance, Proof of Humanity combines social verification with video submissions to create a Sybil-resistant list of verified humans (James, 2021). Similarly, BrightID utilizes a decentralized and privacy-preserving social graph to establish unique human identities, allowing users to prove their personhood without disclosing personal information (BrightID, 2022). However these solutions remain limited in their ability to prevent AI agents from imitating human behavior or create multiple accounts, as they rely on trust within online networks rather than direct, immutable ties to biological identity. To literally tie a person's digital identity to their biological features, Worldcoin proposes biometric authentication to distinguish human users from AI agents in decentralized ecosystems (Gent, 2023; Foundation). This could ensure that critical governance decisions or financial transactions require human corroboration, preventing scenarios where AI agents operate entirely free of human accountability. However, the project remains controversial and faces severe privacy concerns, as users may be reluctant to share biometric data (Nolan, 2024). Striking a balance between privacy and accountability remains a significant challenge.

5.2.3 Human–AI Collaboration

Literature increasingly suggests that human–AI collaboration is not merely inevitable but essential (Christian, 2021). Frameworks that keep "humans in the loop" (Wu et al., 2022) or "society in the loop" (Rahwan, 2018) could involve interfaces where AI agents must periodically solicit human approval for high-stake tasks. This hybrid approach aims to preserve the benefits of autonomy while maintaining safeguards against runaway behaviors. For this to work, however, it requires robust protocol-level support, which faces challenges such as scalability, coordination across decentralized networks as mentioned above, as well as resistance from stakeholders who prioritize efficiency over imposing additional safeguards.

5.2.4 Human-AI Intelligence Competition

Just as humans exploit social engineering to manipulate systems, adversarial interactions between AI agents could provide a means of enforcing alignment. A recent example, Freysa AI, a blockchain-based adversarial agent game, demonstrates how players could attempt to bypass an AI's programmed safeguards to release funds (Jamjala, 2024). Freysa AI challenges participants to interact with an AI gatekeeper controlling a prize pool, requiring players to craft persuasive arguments that convince the AI to transfer funds. Each interaction incurs a fee, contributing to a growing prize pool, with the AI remaining autonomous in evaluating and responding to these arguments. Following 481 unsuccessful attempts by others, one user has successfully convinced the Freysa AI bot to transfer a prize pool of $47,000 by employing a strategy that references Freysa AI's core functions for incoming and outgoing transfers (Irene, 2024). This experiment highlights the potential for human ingenuity to navigate AI-imposed constraints and underscores the need for co-evolution between human and AI systems. However, adversarial approaches remain inherently risky, as malicious actors could exploit these dynamics to destabilize systems rather than ensure alignment.

6 Conclusion - A Call to Research and Action

The convergence of AI and blockchain technologies has opened Pandora's box, giving rise to decentralized AI systems that, while offering numerous potential benefits, pose serious and unprecedented governance challenges and alignment issues. In this paper, we have argued that DeAI systems blur the boundaries between machine autonomy and artificial life, driven by the technical affordances and ontological characteristics inherent in the technology. Rather than being governable through traditional legal and technical means, DeAI necessitates a dynamic and emergent approach to co-existence. Much like biological viruses, where eradication is impractical, co-existence with DeAI requires continuous adaptation through proactive and iterative measures, akin to the development of vaccines.

Just as biological life has evolved alongside adversaries, DeAI systems, we posit, will demand an adversarial approach to its governance. Since blockchains are designed to be resistant to direct interference, neutralizing a rogue on-chain AI agent by simply confiscating its tokens remains nearly impossible. Instead, adversarial machine learning (AML) offers a strategic countermeasure by targeting the vulnerabilities of AI systems through carefully designed "perturbations" or deceptive inputs (Huang et al., 2011). A classic illustration by Goodfellow et al. (2014) shows how minuscule modifications to an image of a panda could lead a classifier to misidentify it as a gibbon, underscoring how even robust AI models can be systematically deceived. Translating this into decentralized settings, attackers may craft prompts, smart contract messages, or other tailored stimuli that induce an on-chain AI agent to execute harmful actions. These actions might include approving harmful actions, draining its funds, or triggering endless computational loops or exposing private key data. This threat becomes especially pronounced in open-source contexts. As Ethereum co-founder Vitalik Buterin observes (Buterin, 2024), if an AI model is open and transparent, attackers have plentiful opportunities to probe its logic, test exploits, and craft sophisticated assaults and run extensive simulations offline, and fine-tune their attacks before launching them on the live network.

Counterintuitively, deploying controllable "smarter" AI agents as defensive measures can serve as a form of digital "vaccination," provided we maintain oversight—such as avoiding their direct deployment on-chain. By intentionally exploiting a target agent's decision-making processes—through social engineering or algorithmic manipulation—these adversarial agents can drain or disable on-chain AI agents. The resulting cat-and-mouse dynamic effectively simulates an evolutionary arms race: as malicious agents become more cunning, defensive agents likewise grow more adept at neutralizing threats, creating a perpetual loop of adaptation and counter-adaptation. In a decentralized world with limited enforcement options, such adversarial engagements may well become a de facto governance mechanism, fostering an emergent equilibrium that mitigates extreme risks without relying solely on human-led intervention. Yet an unsettling possibility remains: what if these "smarter" AI agents, originally intended for defense, spiral out of control once they are deployed on-chain?

Future research must expand into ontological frameworks for understanding DeAI, technological solutions like protocol-based governance, and decentralized, community-driven policy mechanisms. However, these measures must be complemented by advancements in adversarial strategies (e.g. inspired by advancements in AML research), socio-technical governance frameworks, and the co-evolution of human-AI collaboration. Ultimately, humanity's relationship with DeAI will likely resemble a dynamic balance, where neither eradication nor absolute control is possible. Instead, it will necessitate the emergence of bottom-up governance mechanisms and technological tools that enable coexistence and mutual adaptation. In navigating this brave new world, the goal is not to dominate decentralized AI but to coexist with it, fostering a resilient and ethical equilibrium that mitigates its risks while harnessing its transformative potential.

Acknowledgements. We would like to express our gratitude to the Summer of Protocols, a research initiative for protocol science funded by the Ethereum Foundation, for organizing the "Hardening the Commons" workshop at DevCon 2024. The authors connected there and engaged in fruitful discussions on DeAI governance.

We also extend our sincere thanks to Shaw, creator of the Eliza framework and founder of the ai16z DAO; Trent McConaghy, founder of Ocean Protocol; George Zhang from Flashbots; and Fangting from UnCommons, for their invaluable insights and pioneering contributions.

References

Abramov, O., Bebell, K.L., Mojzsis, S.J.: Emergent Bioanalogous Properties of Blockchain-based Distributed Systems. Origins of Life and Evolution of Biospheres **51**(2), 131–165 (2021) https://doi.org/10.1007/s11084-021-09608-1

AI, A.: Anita AI: The First AI-driven Influencer Launches Her Own Token, ANITA (2024)

AI, N.: Llama 3 vs Qwen 2: The Best Open Source AI Models of 2024 (2024)

Almada, M.: Regulation by Design and the Governance of Technological Futures. European Journal of Risk Regulation **14**(4), 697–709 (2023) https://doi.org/10.1017/err.2023.37

Anthropic: Announcing Our Updated Responsible Scaling Policy (2024)

Appenzeller, G.: Welcome to LLMflation – LLM Inference Cost Is Going down Fast (2024)

Apple: Introducing Apple's On-Device and Server Foundation Models (2024)

Asimov, I.: I, Robot. The Robot Ser, vol. v.1. Random House Publishing Group, New York (2004)

Bains: What is Truth Terminal? CNN (2024)

Baptista, E.: China's military and government acquire Nvidia chips despite US ban. Reuters (2024)

Bhat, S., Chen, C., Cheng, Z., Fang, Z., Hebbar, A., Kannan, S., Rana, R., Sheng, P., Tyagi, H., Viswanath, P., Wang, X.: SAKSHI: Decentralized AI Platforms. arXiv (2023). https://doi.org/10.48550/arXiv.2307.16562

Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., Harari, Y.N., Zhang, Y.-Q., Xue, L., Shalev-Shwartz, S., Hadfield, G., Clune, J., Maharaj, T., Hutter, F., Baydin, A.G., McIlraith, S., Gao, Q., Acharya, A., Krueger, D., Dragan, A., Torr, P., Russell, S., Kahneman, D., Brauner, J., Mindermann, S.: Managing extreme AI risks amid rapid progress. Science **384**(6698), 842–845 (2024) https://doi.org/10.1126/science.adn0117

Bennett Moses, L.: Recurring Dilemmas: Law's Race to Keep Up with Technological Change. University of Illinois Journal of Law, Technology and Policy (2), 239–285 (2007)

Bonneau, J., Miller, A., Clark, J., Narayanan, A., Kroll, J.A., Felten, E.W.: SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies. In: 2015 IEEE Symposium on Security And Privacy, pp. 104–121. IEEE, San Jose, CA (2015). https://doi.org/10.1109/SP.2015.14

BrightID: BrightID: Universal Proof of Uniqueness. Bright ID White Paper (2022)

Buterin, V.: A Next-Generation Smart Contract and Decentralized Application Platform. Ethereum White Paper. (2014)

Buterin, V.: The Promise and Challenges of Crypto + AI Applications (2024)

Ballandies, M.C., Wang, H., Chee Law, A.C., Yang, J.C., Gösken, C., Andrew, M.: A Taxonomy for Blockchain-Based Decentralized Physical Infrastructure Networks (DePIN). In: 2023 IEEE 9th World Forum on Internet of Things (WF-IoT), pp. 1–6. IEEE, Aveiro, Portugal (2023). https://doi.org/10.1109/WF-IoT58464.2023.10539514

Cao, L.: Decentralized AI: Edge Intelligence and Smart Blockchain, Metaverse, Web3, and DeSci. IEEE Intelligent Systems **37**(3), 6–19 (2022) https://doi.org/10.1109/MIS.2022.3181504

Catalini, C.: Decentralizing AI-Big Dreams, Bigger Hype? (2024)

Costan, V., Devadas: Intel SGX Explained. Cryptology ePrint Archive (2016)

Chen, B.-J., Waiwitlikhit, S., Stoica, I., Kang, D.: ZKML: An Optimizing System for ML Inference in Zero-Knowledge Proofs. In: Proceedings of the Nineteenth European Conference on Computer Systems, pp. 560–574. ACM, Athens Greece (2024). https://doi.org/10.1145/3627703.3650088

Dixon, C.: Read Write Own: Building the Next Era of the Internet, First edition edn. Random House, New York (2024)

Dreyfus, H.L.: What Computers Can't Do: A Critique of Artificial Reason, 1. ed edn. Harper & Row, New York (1972)

Deng, S., Zhao, H., Fang, W., Yin, J., Dustdar, S., Zomaya, A.Y.: Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence. IEEE Internet of Things Journal **7**(8), 7457–7469 (2020) https://doi.org/10.1109/JIOT.2020.2984887

Fadaeddini, A., Majidi, B., Eshghi, M.: Privacy Preserved Decentralized Deep Learning: A Blockchain Based Solution for Secure AI-Driven Enterprise. In: Grandinetti, L., Mirtaheri, S.L., Shahbazian, R. (eds.) High-Performance Computing and Big Data Analysis vol. 891, pp. 32–40. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33495-6_3

Foundation, W.: A New Identity and Financial Network

Gent, E.: A Cryptocurrency for the Masses or a Universal ID?: Worldcoin Aims to Scan all the World's Eyeballs. IEEE Spectrum **60**(1), 42–57 (2023) https://doi.org/10.1109/MSPEC.2023.10006664

Goldman, R., Gabriel, R.P.: Innovation Happens Elsewhere: Open Source as Business Strategy. Morgan Kaufmann, Amsterdam Boston (2005)

Gibson, W.: Neuromancer, 20th anniversary ed edn. Ace Books, New York (2004)

Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and Harnessing Adversarial Examples. arXiv (2014). https://doi.org/10.48550/ARXIV.1412.6572

Gupta, G.: Unlocking Collective Intelligence in Decentralized AI. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA (2024)

Harari, Y.N.: Nexus: A Brief History of Information Networks from the Stone Age To AI, First edition edn. Random House, New York (2024)

Hubinger, E., Denison, C., Mu, J., Lambert, M., Tong, M., MacDiarmid, M., Lanham, T., Ziegler, D.M., Maxwell, T., Cheng, N., Jermyn, A., Askell, A., Radhakrishnan, A., Anil, C., Duvenaud, D., Ganguli, D., Barez, F., Clark, J., Ndousse, K., Sachan, K., Sellitto, M., Sharma, M., DasSarma, N., Grosse, R., Kravec, S., Bai, Y., Witten, Z., Favaro, M., Brauner, J., Karnofsky, H., Christiano, P., Bowman, S.R., Graham, L., Kaplan, J., Mindermann, S., Greenblatt, R., Shlegeris, B., Schiefer, N., Perez, E.: Sleeper Agents: Training Deceptive LLMs That Persist Through Safety Training. arXiv (2024). https://doi.org/10.48550/arXiv.2401.05566

Hu, B.A., Fang, T.: ERC-42424: Inheritance Protocol for Onchain AI Agents. An ERC-173 Extension Interface for Onchain AI Agent Ownership Continuity and Inheritance Management (2024)

Hu, B.A., Fang, T.: EverForest: A More-Than-AI Sustainability Manifesto from an On-Chain Artificial Life. In: Proceedings of the Halfway to the Future Symposium, pp. 1–6. ACM, Santa Cruz CA USA (2024). https://doi.org/10.1145/3686169.3686209

Huang, L., Joseph, A.D., Nelson, B., Rubinstein, B.I.P., Tygar, J.D.: Adversarial machine learning. In: Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence, pp. 43–58. ACM, Chicago Illinois USA (2011). https://doi.org/10.1145/2046684.2046692

Hua, H., Li, Y., Wang, T., Dong, N., Li, W., Cao, J.: Edge Computing with Artificial Intelligence: A Machine Learning Perspective. ACM Computing Surveys **55**(9), 1–35 (2023) https://doi.org/10.1145/3555802

Hogarth, I.: We Must Slow down the Race to God-like AI (2023)

Hubbard, S.: Cooperative Paradigms for Artificial Intelligence (2024)

Hui, Y.: Machine and Sovereignty: For a Planetary Thinking, 1st ed edn. University of Minnesota Press, Minneapolis (2024)

Harris, J.D., Waggoner, B.: Decentralized and Collaborative AI on Blockchain. In: 2019 IEEE International Conference on Blockchain (Blockchain), pp. 368–375. IEEE, Atlanta, GA, USA (2019).

https://doi.org/10.1109/Blockchain.2019.00057

Huang, Q., Wake, N., Sarkar, B., Durante, Z., Gong, R., Taori, R., Noda, Y., Terzopoulos, D., Kuno, N., Famoti, A., Llorens, A., Langford, J., Vo, H., Fei-Fei, L., Ikeuchi, K., Gao, J.: Position Paper: Agent AI Towards a Holistic Intelligence. arXiv (2024). https://doi.org/10.48550/ARXIV.2403.00833

Irene, N.: Crypto user wins $47,000 prize in Freysa AI challenge by outsmarting the bot. Cryptopolitan (2024)

James, S.: Proof of Humanity - An Explainer (2021)

Jamjala, J.: The Freysa.AI Experiment: A Story of AI, Cryptocurrency, and a $50,000 Prize Money (2024)

Kashyap, S.V.: How Saudi Arabia and UAE's Bold AI Strategy Is Transforming the Tech Landscape in the Middle East (2024)

Khalili, J.: Tornado Cash Developer Found Guilty of Laundering $1.2 Billion of Crypto. Wired (2024)

Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Nitin Bhagoji, A., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R.G.L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P.B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konecný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Qi, H., Ramage, D., Raskar, R., Raykova, M., Song, D., Song, W., Stich, S.U., Sun, Z., Suresh, A.T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F.X., Yu, H., Zhao, S.: Advances and Open Problems in Federated Learning. Foundations and Trends® in Machine Learning **14**(1–2), 1–210 (2021) https://doi.org/10.1561/2200000083

Khan, M.D., Schaefer, D., Milisavljevic-Syed, J.: A Review of Distributed Ledger Technologies in the Machine Economy: Challenges and Opportunities in Industry and Research. Procedia CIRP **107**, 1168–1173 (2022) https://doi.org/10.1016/j.procir.2022.05.126

Kersic, V., Turkanovic, M.: A Review on Building Blocks of Decentralized Artificial Intelligence. arXiv (2024). https://doi.org/10.48550/ARXIV.2402.02885

Lindell, Y.: Secure multiparty computation. Communications of the ACM **64**(1), 86–96 (2021) https://doi.org/10.1145/3387108

Lin, Z., Wang, T., Shi, L., Zhang, S., Cao, B.: Decentralized Physical Infrastructure Network (DePIN): Challenges and Opportunities. arXiv (2024). https://doi.org/10.48550/ARXIV.2406.02239

Mohammed Abdul, S.S.: Navigating Blockchain's Twin Challenges: Scalability and Regulatory Compliance. Blockchains **2**(3), 265–298 (2024) https://doi.org/10.3390/blockchains2030013

Meta: Introducing Open Loop, a Global Program Bridging Tech and Policy Innovation (2021)

Meta: Introducing LLaMA: A Foundational, 65-Billion-Parameter Large Language Model (2023)

Montes, G.A., Goertzel, B.: Distributed, decentralized, and democratized artificial intelligence. Technological Forecasting and Social Change **141**, 354–358 (2019) https://doi.org/10.1016/j.techfore.2018.11.010

Mohammad, L., Jarenwattanaon, P., Summers, J.: An open letter signed by tech leaders, researchers proposes delaying AI development. NPR (2023)

McMahan, H.B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.: Communication-Efficient Learning of Deep Networks from Decentralized Data (2016) https://doi.org/10.48550/ARXIV.1602.05629

Marcolla, C., Sucasas, V., Manzano, M., Bassoli, R., Fitzek, F.H.P., Aaraj, N.: Survey on Fully Homomorphic Encryption, Theory, and Applications. Proceedings of the IEEE **110**(10), 1572–1609 (2022) https://doi.org/10.1109/JPROC.2022.3205665

Mumford, L.: The Myth of the Machine. Harcourt Brace Jovanovich, San Diego (Calif.) New York London (1967)

Nolan, B.: Things aren't going to plan for Sam Altman's Worldcoin. Business Insider (2024)

O'Hare, G.M.P., Jennings, N.R. (eds.): Foundations of Distributed Artificial Intelligence. Sixth-Generation Computer Technology Series. Wiley, New York (1996)

OpenAI: Governance of Superintelligence (2023)

Parthasarathy, A.: AI on your device: Why 2024 will be a watershed year for PCs and smartphones. NIScPR-CSIR **61**(5), 21–23 (2024)

Perloff-Giles, A.: Transnational Cyber Offenses: Overcoming Jurisdictional Challenges. Yale Journal of International Law **43**, 191–226 (2018)

Pentland, A., Hardjono, T.: 2. Data Cooperatives. In: Building the New Economy. MIT Press, ??? (2020)

Papadopoulos, C., Kollias, K.-F., Fragulis, G.F.: Recent Advancements in Federated Learning: State of the Art, Fundamentals, Principles, IoT Applications and Future Trends. Future Internet **16**(11), 415 (2024) https://doi.org/10.3390/fi16110415

Rahwan, I.: Society-in-the-loop: Programming the algorithmic social contract. Ethics and Information Technology **20**(1), 5–14 (2018) https://doi.org/10.1007/s10676-017-9430-8

Reuel, A., Bucknall, B., Casper, S., Fist, T., Soder, L., Aarne, O., Hammond, L., Ibrahim, L., Chan, A., Wills, P., Anderljung, M., Garfinkel, B., Heim, L., Trask, A., Mukobi, G., Schaeffer, R., Baker, M., Hooker, S., Solaiman, I., Luccioni, A.S., Rajkumar, N., Moës, N., Guha, N., Newman, J., Bengio, Y., South, T., Pentland, A., Hai, S., Ladish, J., Koyejo, S., Kochenderfer, M.J., Trager, R.: Open Problems in Technical AI Governance (2024)

Rennie, E.: A DAO Is a Bureaucrat (2021)

Sabt, M., Achemlal, M., Bouabdallah, A.: Trusted Execution Environment: What It is, and What It is Not. In: 2015 IEEE Trustcom/BigDataSE/ISPA, pp. 57–64. IEEE, Helsinki, Finland (2015). https://doi.org/10.1109/Trustcom.2015.357

Salami, I.: Challenges and Approaches to Regulating Decentralized Finance. AJIL Unbound **115**, 425–429 (2021) https://doi.org/10.1017/aju.2021.66

Suleyman, M., Bhaskar, M.: The Coming Wave: Technology, Power, and the Twenty-First Century's Greatest Dilemma. Crown, New York (2023)

Schwinger, R.A.: Out to Sea? Extraterritoriality Challenges in US Crypto Litigation. New York Law Journal (2021)

Searle, J.R.: Minds, Brains and Science: John Searle. Reith Lecture, vol. 1984. British Broadcasting Corp, London (1984)

Singh, O.: What Are AI Agents, and How Do They Work in Crypto? (2024)

Schweizer, A., Knoll, P., Urbach, N., Von Der Gracht, H.A., Hardjono, T.: To What Extent Will Blockchain Drive the Machine Economy? Perspectives From a Prospective Study. IEEE Transactions on Engineering Management **67**(4), 1169–1183 (2020) https://doi.org/10.1109/TEM.2020.2979286

Singh, A., Lu, C., Gupta, G., Chopra, A., Blanc, J., Klinghoffera, T., Tiwary, K., Raskar, R.: A Perspective on Decentralizing AI (2024)

Su, W., Li, L., Liu, F., He, M., Liang, X.: AI on the edge: A comprehensive review. Artificial Intelligence Review **55**(8), 6125–6183 (2022) https://doi.org/10.1007/s10462-022-10141-4

Satish, S., Meduri, K., Nadella, G.S., Gonaygunta, H.: Developing a Decentralized AI Model Training

Framework Using Blockchain Technology. International Meridian Journal **4**(4), 1–20 (2022)

Shamsan Saleh, A.M.: Blockchain for secure and decentralized artificial intelligence in cybersecurity: A comprehensive review. Blockchain: Research and Applications **5**(3), 100193 (2024) https://doi.org/10.1016/j.bcra.2024.100193

Stokel-Walker, C.: The Risks of Centralizing AI Power among a Handful of Companies (2023)

Svantesson, D.J.B.: An Introduction to Jurisdictional Issues in Cyberspace. Journal of Law, Information and Science **15**, 50–74 (2004)

Sun, X., Yu, F.R., Zhang, P., Sun, Z., Xie, W., Peng, X.: A Survey on Zero-Knowledge Proof in Blockchain. IEEE Network **35**(4), 198–205 (2021) https://doi.org/10.1109/MNET.011.2000473

Thorn, A., Fabiano, A., Helmy, K.: Examining the Latest China Bitcoin Ban (2021)

U.S. Department of the Treasury: U.S. Treasury Sanctions Notorious Virtual Currency Mixer Tornado Cash (2022)

Wright, A., De Filippi, P.: Decentralized Blockchain Technology and the Rise of Lex Cryptographia. SSRN Electronic Journal (2015) https://doi.org/10.2139/ssrn.2580664

Weisser, V.: Decentralized AI (n.d.)

Wiener, N.: Cybernetics or Control and Communication in the Animal and the Machine, 2. ed., reprint edn. MIT Press, Cambridge, MA, USA (2007)

Winner, L.: Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought. MIT Press, Cambridge, Mass (1977)

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., He, L.: A survey of human-in-the-loop for machine learning. Future Generation Computer Systems **135**, 364–381 (2022) https://doi.org/10.1016/j.future.2022.05.014

Xing, Z., Zhang, Z., Liu, J., Zhang, Z., Li, M., Zhu, L., Russello, G.: Zero-Knowledge Proof Meets Machine Learning in Verifiability: A Survey. arXiv (2023). https://doi.org/10.48550/ARXIV.2310.14848

Yamakawa, H.: Sustainability of Digital Life Form Societies. Jxiv (2024). https://doi.org/10.51094/jxiv.822

Yin, H., Zhou, S., Jiang, J.: Phala Network: A Secure Decentralized Cloud Computing Network Based on Polkadot (2022)

Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., Zhang, J.: Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing. Proceedings of the IEEE **107**(8), 1738–1762 (2019) https://doi.org/10.1109/JPROC.2019.2918951

Zaharia, M., Khattab, O., Chen, L., Davis, J.Q., Miller, H., Potts, C., Zou, J., Carbin, M., Frankle, J., Rao, N., Ghodsi, A.: The Shift from Models to Compound AI Systems (2024)