INATBA REPORT

Report on Artificial Intelligence and Blockchain Convergences

JULY 2024

AI & Blockchain Convergences Task Force

In collaboration with

INATBA
International Association for Trusted Blockchain Applications

EUROPEAN BLOCKCHAIN ASSOCIATION

AI and blockchain convergence

Authors	Mariana de la Roche (Validvent - INATBA- tBt) Erwin Voloder (EBA)
Contributors	Ankur Banerjee (CTO, Cheqd) Clara Guerra (Stabsstelle Finanzplatz Innovation & Digitalisierung - Office for Financial Market Innovation) Dino cataldo Dell'Accio (UNJSPF) Fabio Budris (Government of Buenos Aires) Gladstone M. Arantes Jr (BNDES, Brazil) Javed Khattak (Cheqd, FIA) Kai-Ti Wu (PhD Candidate Humboldt Universität) Laura Kajtazi (Validvent) Paolo Giudici (University of Pavia) Stefania Tonutti (Vechain) Tan Gürpinar (Quinnipiac University) Tomaz Sedej (Hyperledger Foundation, Copenhagen Business School)
Reviewers	Ana Felicitas, (Universidad Rey Juan Carlos) Daniel Schoenberger (Web3 Foundation: Polkadot, Kusama) Idris Demir (Batman University, Türkiye) Katarina Krüger (Adam) (Hochschule für Technik und Wirtschaft Berlin - HTW Berlin) Jacques Bughin (ValueVerse) Luis Carro (University of Valladolid) Mathew Yarger (Demia) Nena Dokuzov (Slovenia Government) Sebastian Becker (INATBA) Ximena Puente (World Bank)

1. Introduction **3**
2. Overview of AI and Blockchain Convergence **5**
2.1. Enabling Dataconomy 5
2.2. Central AI vs Decentralized AI 6
2.3. Using blockchain to leverage decentralized training for AI models 7
3. Case examples on Real-World Applications with AI and Blockchain **8**
3.1. AI and decentralized finance (DeFi) 9
3.1.1. AI-Driven trading strategies 9
3.1.2. Fraud detection and security 10
3.1.3. Privacy and AI in DeFi 11
3.1.4. Natural Language Processing (NLP) for Data Analysis 12
3.2. Enabling Digital Twins 13
3.3. Bringing the Metaverse closer to reality 14
3.4. Sustainability & Authenticity 15
3.5. Health Applications 16
4. Challenges and Solutions **16**
4.1. Data Privacy, Ownership, and Blockchain 17
4.2. Using blockchain to secure data sharing and access control 21
4.3. Using blockchain to improve data provenance and audit trails 23
4.4. Using blockchain to combat disinformation 24
4.5. Using blockchain to improve consent management with AI models 25
4.6. Using blockchain to improve data retention and deletion with AI 25
4.7. Combining blockchain and AI to improve data marketplace and incentives 26
5. Ethical Social and Governance (ESG) Considerations **34**
5.1. Governance and Voting Assistance: 34
5.2. Trust and Adoption 35
5.3. Education and Literacy 36
5.4. Acceptable economic models 37
5.5. Global Collaboration 40
6. Conclusion: The Road Ahead: Future Prospects **43**
7. Bibliography **44**

Glossary:

AI	Artificial Intelligence
AMM	Automated Market Maker
B2C	Business-to-Consumer
BIS	Bank for International Settlements
C2C	Consumer-to-Consumer
CVaR	Compounded Value-at-Risk
DApp	Decentralised Application
DAO	Decentralised Autonomous Organisation
DeFi	Decentralised Finance
DEX	Decentralised Exchange
DPos	Delegated Proof-of-Stake
ECOSOC	United Nations Economic and Social Council
EU	European Union
EIC	European Innovation Council
FHE	Fully Homomorphic Encryption
fhEVM	Fully Homomorphic Ethereum Virtual Machine
G20	Group of 20
GAFAM	Google, Apple, Facebook, Amazon, Microsoft
GDPR	General Data Protection Regulation
GenAI	Generative Artificial intelligence
IEC	International Electrotechnical Commission
IGF	Internet Governance Forum
ISO	International Standards Organisation
ITU	International Telecommunications Union
LLM	Large Language Model
LP	Liquidity Provider
MPC	Multiparty Computation
NLP	Natural Language Processing
PBFT	Practical Byzantine-Fault Tolerance
PHE	Partial Homomorphic Encryption
PoS	Proof-of-Stake
PoW	Proof-of-Work
SBT	Soulbound Token
SDG	Sustainable Development Goals
SSI	Self-Sovereign Identity
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organisation
VaR	Value-at-Risk
VCs	Verifiable Credentials
zk-SNARK	Zero-Knowledge Succinct Non-Argument of Knowledge
Zk-STARK	Zero-Knowledge Scalable Transparent Argument of Knowledge

1. Introduction

In the digital age, data has emerged as the cornerstone for innovation. However, this surge in data-driven innovation is not without its challenges. Concerns about user confidentiality, the potential misuse of personal information, and the ever-present risk of breaches are increasingly being highlighted in global discourse (Gelhaar, et al. 2021; Greenwood, S. 2016). Additionally, our interconnected digital ecosystems have exacerbated the rise of misinformation and fake news, posing significant threats to informed decision-making and societal trust (Stanford UIT 2024; Allcott, H., & Gentzkow, M. 2017). The role of AI, particularly with the potential of promising emerging technologies, such as the deep fake technology must be mentioned. (Tyagi, 2023a) In light of the promising effects of deep fakes (such as in marketing and advertising, and the healthcare), these technologies can not be outrightly banned, as also underscored in the recent EU AI Act (European Commission, 2024). Interestingly, the convergence of Distributed Ledger technologies, that is using blockchain to detect the source of deep-faked images and videos can help segregate authentic from the inauthentic content (Tyagi, 2023a).

Before delving into the convergence of AI and blockchain technologies, it's important to recognize their independent advancements and the specific domains they are transforming. Artificial Intelligence has progressed from theoretical concepts to practical, impactful applications. Major strides in machine learning, particularly in deep learning, have enabled AI to process and analyze data at unprecedented speeds and accuracy, leading to innovations in fields ranging from autonomous driving to personalized medicine. Meanwhile, blockchain technology has matured beyond its initial cryptocurrency applications to become a fundamental tool in enhancing security, transparency, and efficiency across various industries. It has played a key role in redefining supply chain management, financial services, and secure digital transactions. These separate advancements set the stage for a transformative synergy when AI and blockchain technologies are integrated.

Historically, there was a delineation between traditional software, rooted in static datasets and established rules, and the expanding world of data analytics, encompassing advanced data science and reinforcement learning—what we now umbrella under the term ‘artificial intelligence’ or 'AI' (Russell & Norvig, 2020). At its core, AI is about deriving insights from data, learning from it, and subsequently, making informed decisions (OECD, 2023). As the usage of AI spreads across varied sectors, ensuring the credibility of its decisions is necessary, not only for operational excellence but also for upholding public trust and ethical standards (Borenstein & Howard, 2020). Today we count different classes of AI. During the last years there have been significant advancements in artificial intelligence (AI), particularly in deep learning, enabling the learning from existing data and the creation of new designs (Liao et al., 2024). Generative AI, which involves the use of machine learning algorithms to learn from data and produce new content, has been identified as a leading technological trend (Alloghani, 2024). Particularly examples such as ChatGPT developed by OpenAI, and AlphaFold by DeepMind have showcased the adaptability of generative AI across various fields (Liao et al., 2024). Today's advanced generative AI products, with their vast capabilities, offer numerous benefits, automating and simplifying traditional tasks. However, the

methodologies these AI models employ have raised questions about the reliability of data sources, the quality of the information produced, and the ownership of this access (Russell & Norvig, 2020).

Blockchain, characterized by its decentralized nature, immutability, and heightened security, stands out as a transformative technology in this landscape (Narayanan et al., 2016). It promises to elevate transparency and potentially reduce or even eliminate the role of intermediaries. However, the journey of blockchain adoption is laden with challenges: from issues of scalability and the complexities of integration with existing systems to the intricate dance of navigating the regulatory landscape (Gürpinar et al. 2024).

AI, with its ability to mimic learning, reasoning, and adaptability, presents itself as a powerhouse of modern computation. Its scope spans from the simplicity of automation to the intricacies of complex decision-making (Russell & Norvig, 2020). Yet, implementing AI is not without its challenges. The sheer volume of data it demands brings forth serious privacy concerns (Zuboff, 2019). Additionally, its often opaque decisions can inadvertently perpetuate biases or be based on inaccurate or even misleading data (Buolamwini & Gebru, 2018); (Babaei, Giudici & Raffinetti, 2024); (Giudici & Raffinetti, 2023). These, coupled with ongoing concerns about its security and scalability (Goodfellow et al., 2020), highlight the urgent need for meticulous regulatory and ethical considerations.

Blockchain's inherent properties might offer solutions to some of AI's challenges. Its decentralized architecture can enhance the diversification of AI data sources, reducing inadvertent biases in AI outputs. A significant advantage of blockchain is its ability to integrate incentive systems, letting people share data in a controlled way and earn rewards (Tapscott & Tapscott, 2016). This empowers individuals by giving them more control over their data, signaling a change in traditional data collection methods. Moreover, blockchain's immutable nature ensures a transparent, traceable, and unchangeable record of AI decisions (Narayanan et al., 2016), promoting trust and accountability in AI systems.

On the other side with its prowess in handling vast datasets and pattern recognition, AI can streamline and enhance blockchain's scalability, detect and rectify anomalous behaviors, and potentially prevent hacks, money laundering, and other illegal activities. Moreover, AI's user-friendly interfaces can simplify the complexity of blockchain interactions for everyday users, driving broader adoption and even being used for educational purposes. AI could also enhance the flexibility of smart contracts. Traditionally, smart contracts are limited to executing predefined algorithms, requiring every possible outcome to be anticipated and coded in advance. This approach results in a rigid functionality. Integrating AI capabilities could enable smart contracts to adapt to a wider range of situations dynamically. Some proposals seem to go in a similar direction (ACM paper, SingularityNET, Giza, EZKL). However, this also significantly increases the importance of ensuring that the AI systems are trained on appropriate data, thereby fostering a virtuous cycle of improvement between the two technologies (EUBOF, 2020, Galaxy Research).

As we delve into the potential of AI and Blockchain, this paper will explore the ways these technologies can converge to redefine industries and societal norms. We will explore their role in enabling data economies, reshaping everything from decentralized finance to healthcare, and advancing the sustainability and authenticity of digital interactions. This exploration will cover topics ranging from AI-driven trading strategies in DeFi to enhancing digital twins, from reimagining the metaverse to fostering community-based innovation.

Furthermore, we address the challenges and propose solutions for securing data sharing, enhancing data provenance, and leveraging blockchain for improved data integrity and consent management. By fostering an ethical framework and encouraging global collaboration, we aim to guide these technologies towards a future where they contribute positively to society, ensuring they enhance rather than compromise our ethical and governance standards. Through this comprehensive analysis, we advocate for a unified approach that views technology not in silos but as an integrated framework aimed at driving a sustainable, equitable, and technologically empowered future.

2. Overview of AI and Blockchain Convergence

In today's digitized world, both AI and blockchain stand as beacons of transformation, each offering profound shifts in their respective domains (Hussain & Al-Turjman, 2021). Combining these two technologies not only augments their strengths but also paves the way for a dynamic synergy (Dinh & Thai, 2018). In this confluence, AI can thrive with enhanced trust, transparency, and efficiency, while blockchain systems stand ready to benefit in terms of optimizing operations and increasing security when it comes to prevention and faster detection of money laundering and financing of terrorism (Dinh & Thai, 2018).

Understanding the individual advancements in AI and blockchain is crucial for appreciating their combined potential. AI's capabilities in data analysis and decision-making, when paired with blockchain's features of decentralization, transparency, and security, can lead to innovative solutions that address current technological and societal challenges. For instance, AI can enhance blockchain's efficiency and security, while blockchain can provide a robust frameworks for data integrity and trust in AI systems. This convergence is poised to revolutionize various sectors, offering new paradigms in data management, automation, and secure, decentralized operations.

2.1. Enabling Dataconomy

Data serves as the cornerstone for AI systems (Russell & Norvig, 2020). The richer the data an AI model receives, the more it can refine its functions (Russell & Norvig, 2020). For instance, many modern text-based generative AI is trained on a dataset called Common Crawl, an archive of a majority of the Web, at 3.35 billion pages worth 450 TB of data. This emphasis on data, however, introduces critical concerns regarding its access, security, authenticity, and overall integrity. This is precisely where blockchain steps in as a solution. By offering a

secure, tamper-proof platform, blockchain could ensure that AI is fed with genuine, authentic and high-quality data as long as the blocks do not contain any malicious data. In practical terms, this does not necessarily mean the training data for AI itself needs to be stored on a blockchain, but could take the form of cryptographic proof of its quality or provenance being secured on a blockchain. It also could clarify the decision-making processes within AI, addressing the often-cited "black box" criticism (Nassar et al., 2019). Furthermore, blockchain encourages a decentralized AI framework, where multiple nodes collectively participate, preventing any single entity from holding absolute control (Montes & Goertzel, 2019).

Through blockchain, a transformative approach can emerge wherein individuals can manage their data, decide on its accessibility, and even monetize it (Montes & Goertzel, 2019). This paints a picture of a future where individuals stand at the center of data ecosystems, with AI systems seeking permissions and possibly offering compensation for user insights and data (Montes & Goertzel, 2019). On the other hand, while blockchain offers much to AI, it too can benefit from AI's capabilities (Dinh & Thai, 2018). Despite its robustness, blockchain is not entirely immune to threats. Here, sophisticated AI algorithms can oversee blockchain network activities, identifying and mitigating anomalies, thereby strengthening blockchain's defenses (Salah et al., 2019).

2.2. Central AI vs Decentralized AI

In exploring the convergence of AI and blockchain, it is crucial to understand the distinct characteristics and implications of centralization in AI and decentralization in blockchain technology. Centralized AI systems, where data and decision-making processes are consolidated by a single entity or set of related entities, offer significant computational power and efficiency. These systems can execute complex algorithms, manage large datasets, and provide rapid insights (Montes & Goertzel, 2019). However, this centralized approach often leads to a concentration of power and control (Cihon et al., 2020), raising concerns over transparency and privacy. Centralized AI systems can inadvertently reinforce biases if they rely on homogenous or skewed datasets, necessitating diverse data sources for comprehensive and unbiased AI decision-making. Integrating various perspectives and data points allows AI to generate more accurate, inclusive, and ethically sound outcomes.

In contrast, decentralizing AI involves diversifying data sources and decision-making processes across multiple points (Demazeau & Müller, 1991). Moreover, decentralized AI goes beyond just distributed data sources and decision-making. It prioritizes open-source principles, fostering collaboration and transparency throughout the AI development lifecycle. This approach aligns with the core characteristics of blockchain, emphasizing distribution and decentralization (Grosse et al. 2020). Moreover, the transparent and independently auditable nature of blockchains makes it easier to inspect whether the datasets used for training are, in fact, diverse and reduce bias. By sourcing data from a diverse set of decentralized nodes, AI systems can access a richer, more varied pool of information, significantly reducing the risk of biases and enhancing system security against typical failures of centralized systems (Cao, 2022).

Decentralized Artificial Intelligence DAI, as explained by Cao, (2022), involves "storing, updating, sharing, and exchanging decentralized intelligence between decentralized agents, nodes, or devices; and integrating decentralized intelligence from local agents and across decentralized ecosystems". This approach includes using blockchains and federated learning to enable autonomous agents on devices to carry out AI tasks locally, either as independent entities or part of a connected network, maintaining privacy and decentralizing decision-making (Cao, 2022). In effect, decentralized AI leverages the strength of varied data inputs, ensuring a more holistic and representative view (Cao, 2022). This diversity is vital in training AI models that are resilient, least prompt to biases, and adaptable to different scenarios and populations (Cao, 2022).

In practice, this means building AI systems that are not only powerful and efficient but also ethically sound, transparent, and inclusive, by building them in a more representative and collective manner (Cao, 2022). The integration of decentralized AI models and blockchain technologies is set to democratize AI development, reducing the dominance of large technology vendors and increasing transparency in AI decision-making (Montes & Goertzel, 2019).

SingularityNET presents a compelling case of decentralized AI in action. By creating an open marketplace for AI services, it empowers developers and businesses to deploy and exchange AI capabilities seamlessly. This decentralized network of AI services, built on blockchain technology, facilitates not only a diversified range of AI solutions but also promotes a cooperative environment where AI systems can learn from each other, thus improving over time. Another great example of decentralized AI is the Ocean Protocol which uses blockchain technology to democratize access to data, a critical resource for AI development. Through its decentralized data marketplace, Ocean Protocol enables a wide range of data providers to contribute to AI training datasets, thereby enhancing the diversity and quality of data available for AI models. This approach not only promotes data availability but also ensures data privacy and user control, aligning with the principles of decentralized AI.

However, integrating these two technologies is not without challenges. From the technical complexity of harmonizing systems that operate under very different principles to concerns about the privacy and governance of data generated or processed by AI on blockchains. Furthermore, scalability remains a crucial issue, as both technologies must be able to handle increasing volumes of operations without compromising their performance or security, METLABS.

2.3. Using blockchain to leverage decentralized training for AI models

In looking at transformer based architectures, blockchain could provide benefits through decentralized training. Traditional centralized training methods often face challenges in efficiently distributing computational workloads across multiple processors or GPUs due to limitations in bandwidth, communication overhead and synchronization bottlenecks. In contrast, blockchain networks offer a decentralized infrastructure where training tasks can

be parallelized and distributed across a vast network of nodes, each contributing computational resources to the training process.

Blockchain's distributed nature allows for the creation of a peer-to-peer network where nodes can collaborate in the training of AI models without relying on a centralized server - enabling the simultaneous execution of training tasks on multiple nodes. This leads to faster convergence and reduced training times for large-scale models. Blockchain's consensus mechanisms ensure that all nodes in the network agree on the validity of training updates, maintaining data integrity and preventing inconsistencies in the model's parameters.

Furthermore, blockchain-based decentralized training facilitates the aggregation of gradients or model updates from different nodes in a secure and transparent manner. Each node independently computes gradients using its local data and contributes these updates to the global model through a consensus mechanism. This parallelized approach enables efficient utilization of computational resources and accelerates the training process for transformer-based models with billions of parameters.
It is also possible to pool computational functions in decentralized training for AI models. For example, Secure Multiparty Computation (MPC) is a cryptographic technique allowing multiple parties to jointly compute a function over their private inputs while keeping those inputs confidential (Knott et al., 2021). With AI models, MPC could be applied to address privacy concerns and maintain the confidentiality of data (Knott et al., 2021). In decentralized training, participants (nodes) could apply MPC to compute model updates based on their local data. During decentralized training, model updates from multiple nodes would also need to be aggregated. MPC can be used to securely aggregate any model updates, ensuring that the final model parameters result from a joint computation rather than exposing individual contributions (Knott et al., 2021). This helps prevent potential information leakage during the aggregation phase, making it suitable for privacy sensitive AI applications (Knott et al., 2021). By applying MPC to AI models, nodes in a decentralized training network could agree on a model update without revealing their individual gradients, enabling secure consensus by combining encrypted inputs from all nodes. Nodes can therefore determine the next model update without revealing their local data. Additionally, open-source MPC libraries, (e.g. SCALE-MAMBA).can help ensure trust and create transparency in decentralized AI training. By allowing public scrutiny of the code, these libraries empower developers to verify the security and proper functioning of MPC implementations.One potential drawback in applying MPC to decentralized models is the computational intensity. AI training models often require complex, large data sets (Daglarli, 2019) therefore employing efficient MPC protocols and improvements in hardware acceleration are crucial to achieving practical performance at scale.

3. Case examples on Real-World Applications with AI and Blockchain

This chapter goes into the practical and transformative applications of AI and blockchain technology across various sectors. It presents the synergy between these two cutting-edge technologies and how they are being leveraged in real-world scenarios, offering innovative

solutions and reshaping industries. From the complexities of DeFi and the intricacies of digital twin technologies to the expansive potential of the metaverse, the intersection of AI and blockchain demonstrates a remarkable capacity for enhancing sustainability, authenticity, and healthcare applications.

3.1. AI and decentralized finance (DeFi)

Automated market making (AMM) is a fundamental component of decentralized exchanges (DEXs) such as Uniswap, Balancer and Curve (Xu et al., 2023). Liquidity providers (LPs) deposit asset into liquidity pools to facilitate trading without relying on traditional order books (Xu et al., 2023). As liquidity provision involves users depositing funds into these pools, AI can play a key role in optimizing the process (Rabetti, 2023).

3.1.1. AI-Driven trading strategies

AI algorithms can analyze vast amounts of historical trading data, order book dynamics, and market sentiment indicators to develop and execute strategies for Automatic Market Makers- AMM – continuously adjusting prices and liquidity provision (LP) (Xu et al., 2023). This can aid LPs in setting more competitive pricing, adjusting their positions in real-time and adapting to changing market conditions (Xu et al., 2023). For example, various machine learning algorithms including regression, time series forecasting, neural networks and ensemble methods can be applied to predict asset price movements. Reinforcement learning can also be used to optimize trading strategies (Alameer, et al. 2022).

Therefore, AI can capture the complex relationships between data to apply, and adapt on-the-fly to changing market conditions (Sullivan & Wamba 2024). AI models can also incorporate sentiment analysis, by monitoring news feeds and social media chatter which in the world of crypto trading is crucial as the industry operates under the umbrella of so-called 'narrative economics (Hamdan et al., 2021).' More broadly, AI can be used to create full portfolio-driven strategies rather than focusing specifically on individual assets. Modern portfolio optimization techniques such as Markowitz's Mean Variance Optimization (Giudici & Polinesi, 2022); (Giudici, Leach & Pagnottoni, 2022), or more advanced techniques such as the Blackman-Litterman model (He & Litterman, 2002). This extends traditional mean-variance optimization by incorporating subject views and forward-looking expectations into the portfolio allocation process (He & Litterman, 2002). It is particularly useful in situations where historical data alone may not fully capture market dynamics, as is often the case in DeFi.

AI can dramatically adjust the prices at which assets are bought and sold within liquidity pools in correspondence with changing supply and demand conditions. AI algorithms can also be deployed to assess the risk associated with specific pools and rebalance positions accordingly. For example, AI-driven risk management models can calculate Value at Risk (VaR) and Conditional Value at Risk (CVaR) (Giudici & Abu-Hashish, 2019); (Ahelegbey & Giudici, 2022), recommending position sizing and stop loss levels. Such tactics are

important in DeFi trading where volatility is high. Dynamic pricing in turn ensures that LPs offer competitive rates, which leads to more traders on a given DEX. Additionally, AI can help LPs manage exposure to more volatile assets and act as a check against impermanent loss. AI algorithms can continuously monitor asset prices, trading volumes and orderbook data (Amirzadeh et al., 2022). This can be very useful for the monitoring and trading across multiple DeFi protocols and DEXs allowing to identify instances when an asset's price is lower on one platform or higher on another.

This can allow users to take advantage of cross-protocol arbitrage, capitalizing on price disparities. Through automated execution, AI-powered trading bots can execute arbitrage trades swiftly, placing orders on different platforms simultaneously to capture price differentials (Sifat, 2023). AI can further analyze liquidity pool data across various DEXs and lending platforms, assessing factors such as pool depth, trading volume, fees, and slippage (Basly, 2024). This helps in identifying pools with the most favorable conditions for liquidity provision (Basly, 2024). In terms of risk management, AI models can help assess the risks associated with arbitrage and liquidity allocation strategies, considering factors such as market volatility, transaction costs and potential liquidity shortages (Sadman et al., 2021). This helps the model recommend appropriate position sizing and risk mitigation strategies as necessary. By leveraging dynamic liquidity allocation, AI models can automatically reallocate assets to pools with more favorable terms or higher yield (Basly, 2024). Through data aggregation and real-time updates, AI systems can aggregate data from multiple sources including DEXs, lending protocols, layer 1/layer 2 blockchains and roll-ups, giving users up-to-date information on market conditions and opportunities (Basly, 2024).

Today we can already observe AI-powered bots such as earlybird, which automatically reinvest earned rewards or staking payouts into additional yield farming opportunities or assets. AI can also be used to adapt yield farming and staking strategies based on changing market conditions (Nartey, 2024) such as fluctuations in gas fees, network congestion or protocol upgrades. For example, the algorithm could set thresholds for gas fees, automatically triggering actions when fees fall within specified ranges. By detecting network congestion and/or transaction backlog and block confirmation times, AI could identify optimal entry points to perform DeFi activities such as providing liquidity or claiming rewards. AI can keep track of protocol upgrades or governance decisions, adjusting staking strategies to align with new rules or opportunities. Moreover, AI algorithms can assess the risk-return profile of different yield farming and staking opportunities and adjust the allocation of assets accordingly through optimal pool selection techniques (Nartey, 2024).

3.1.2. Fraud detection and security

AI algorithms can analyze user behavior, transaction patterns and historical data to establish normal profiles. Any deviations from these profiles such as unusually large transactions or irregular trading activity can trigger alerts for further investigation. This sort of behavioral analysis can identify fraud such as front-running or suspected insider trading. Moreover, machine learning algorithms can be trained to detect anomalies in transaction

data, including any outliers or patterns that do not conform to expected 'norms' (Luo et al., 2023). Anomaly detection can be done through statistical methods or with advanced algorithms like autoencoders. Additionally, AI can perform clustering and link analysis on blockchain addresses to group together related addresses that may be associated with the same entity, further helping track suspicious activity across multiple addresses (Luo et al., 2023). This is important when trying to unravel complex fraud schemes, money laundering or the movement of stolen assets across the blockchain (Giudici & Raffinetti, 2022; Aldasoro et al., 2022).

AI can also be helpful when it comes to pattern recognition in smart contracts, analyzing them to detect vulnerabilities, code exploits and potential security risks (Luo et al., 2023). In DeFi, where hacks and attempted hacks are a growing concern, machine learning algorithms that can identify patterns indicative of malicious code can help secure both the protocol and user funds. Natural language processing (NLP) models can analyze news articles, social media and online forms to gauge market sentiment, identifying discussions related to potential fraud or security threats (Luo et al., 2023). AI powered monitoring systems can be further scaled to handle large volumes of transactions and provide real-time alerts to security teams or users when suspicious activities are detected (Luo et al., 2023).

3.1.3. Privacy and AI in DeFi

It is therefore well established that DeFi is an expanding domain characterized by financial applications executed through smart contracts within the blockchain infrastructure and this has introduced challenges related to utilizing privacy data to enhance market efficiency while safeguarding user privacy (Zhuangtong Huang et al., 2023). In traditional finance, where we have intermediaries, there is granted a "control", a private access and so the privacy of data subjects; in DeFi, since platforms operate on blockchain networks, anyone can access them without any control and privacy, in a total transparency manner. Adding to this, today the combination of DeFi and AI offers several opportunities to enhance the decentralized model, making it even more transparent, where security also needs to be considered: as AI algorithms are susceptible to data manipulation, there could be risks of loss due to inaccurate predictions. That is why it is absolutely necessary to have a risk-based and above all multidisciplinary approach. Today, the so called "Privacy Paradox" in DeFi sets several challenges: a) on one hand the blockchain transactions are pseudonymous (because the users are represented by cryptographic addresses) but the analytical tools could potentially de-anonymize them, posing privacy concerns (Diana Ambolis, 2024); b) the transparent nature of the blockchain means that all transaction are recorded on the-chain and this could exposes data subjects to security and financial risks (Diana Ambolis, 2024); c) the immutable nature of the blockchain forecloses the right to amend and forgot of the data subjects granted by art. 16 - 17 GDPR. An answer to these challenges could be AI: AI -Driven solutions to grant a DeFi Privacy.

Applying federated learning, which is defined as a machine learning approach where a model is trained across multiple decentralized devices or servers holding local data samples without exchanging them (McMahan et al., 2017), to DeFi can be used to analyze transaction

data, trading behaviors, and liquidity provision across multiple users without exposing individual user data. For example, a DeFi platform can use federated learning to gain insights into user preferences for liquidity pool selection without knowing specific user actions. In DeFi lending, secure multiparty computation (MPC) could be used to calculate interest rates and loan terms without revealing the borrower's financial details to lenders or vice versa (Baum et al., 2023). Furthermore, homomorphic encryption[1] can be used on DeFi platforms to perform calculations related to portfolio diversification, risk assessment, or token swaps without exposing user holdings and/or trading strategies.

Zero knowledge proofs such as zk-SNARKs or zk-STARKs provide an alternative mechanism to achieving similar goals as secure MPC and (explained in detail under section 4) can be applied to verify transactions, user balances and smart contract compliance without disclosing transaction details (Baum et al., 2023). For example, a DeFi protocol, can deploy a SNARK proof to verify the validity of a transaction without revealing the sender, receiver or transaction amount. There is also the option to use differential privacy techniques, adding noise to query responses which can help protect user specific data when analyzing market trends, liquidity pool utilization or governance voting patterns. AI can also be incorporated into confidential smart contracts (e.g. fhEVM smart contracts, or Secret Network). For example, when a user wants to make predictions or inferences using an AI model, they can send their encrypted input data to Secret Network, which performs computations on the encrypted data, ensuring it remains private throughout the process. Conversely, AI-services such as prediction models or data analysis tools can be deployed as decentralized applications (Dapps) on fhEVM. This would enable Fully Homomorphic Encryption (FHE) operations on the Ethereum Virtual Machine and facilitate confidential payments for accessing these services.

3.1.4. Natural Language Processing (NLP) for Data Analysis

Sentiment analysis (also known as opinion mining) assesses the sentiment or emotional tone expressed in text data. NLP models classify text as positive, negative or neutral to gauge market sentiment. In DeFi, NLP can monitor social media posts and news articles to gauge sentiment toward specific DeFi projects, cryptoassets or market trends (Cerchiello, Giudici & Nicola, 2017). Automated event detection can also provide real-time updates on DeFi project announcement, protocol upgrades, regulatory developments and market moving news to help traders and investors stay informed. It can also be used in the reverse, to help regulators and policy makers assess the overall health and stability of the DeFi ecosystem by staying on-top of both positive and negative market events.

NLP can also be used to track changes in market sentiment over time (Paramanik & Singhal, 2020), taking stock of how shifts in sentiment towards specific DeFi projects move as

[1] Homomorphic encryption allows computations on ciphertexts, producing encrypted results that match the plaintext operations once decrypted. This feature is crucial for modern communication systems. RSA was the first public-key encryption scheme with a homomorphic property, but it loses this property due to the need for padding messages with random bits for security. To address this, numerous homomorphic encryption schemes have been developed over the past three decades (Yi et al., 2014).

market conditions fluctuate. Entity recognition is also possible through NLP techniques, with NLP models able to identify and categorize entities mentioned in text data including crypto assets, DeFi projects or influential figures in the space. Through event impact analysis, NLP can assess the impact of specific events or news (Yadav et al., 2020) in the DeFi market by analyzing how sentiment and trading volumes respond after the market shock. On a community level, NLP can be used to track community sentiment (Eliaçik & Erdoğan, 2015), allowing DeFi projects to better understand user concerns, preferences and voting patterns which can lead to more informed governance decisions.

3.2. Enabling Digital Twins

The fusion of Generative AI with digital twin technology brings a new era of innovation in different sectors (Bariah & Debbah, 2022) such asset-heavy industries like manufacturing or utilities. This combination facilitates an interesting approach to managing and monitoring critical assets. Using advanced neural network architectures applied to extensive visual data, we can develop sophisticated models for diverse assets such as machinery, electrical systems, and supply chains. This method significantly surpasses traditional inspection techniques, enabling quick detection of anomalies and damages, and offering real-time insights into asset health.

Beyond mere visualization, these models evolve into individual digital twins of assets. By harnessing time series data, alongside work orders and event predictions, these digital twins offer a comprehensive historical view, essential for superior anomaly detection and predictive maintenance. This predictive approach not only enhances asset performance but also extends its operational lifespan, ensuring more efficient and reliable operations.

Additionally, the incorporation of Generative AI in field service support is transforming how support is delivered. Through retrieval-augmented generation tasks, AI can provide real-time Q&A support and multilingual conversational assistance. These AI-driven tools, grounded in extensive knowledge bases, offer immediate and relevant guidance, boosting the efficiency of field service teams. This is a significant step forward in reducing dependency on manual searches or external human support, leading to faster resolution of field issues. However, recently, there have been cases of generative AI hallucinating in support scenarios. Such events are not grounds to exclude the wholesale application of AI in support scenarios, but underscore the need to improve their execution by grounding the chat bot in 'reality' via fine-tuning against actual corporate knowledge bases it is meant to support. The journey of integrating Generative AI and digital twins is not without its challenges. Building trust and ensuring transparency in AI systems, particularly in Generative AI and Large Language Models (LLMs), are crucial hurdles (Nah et al., 2023). Often, AI projects face roadblocks in their proof-of-concept stages due to strategic misalignments or doubts about the model's outcomes. Navigating these challenges calls for more than just technological solutions; it demands a socio-technological approach that recognizes the broader implications and impacts of AI in real-world settings. This approach is key to unlocking the full potential of AI and digital twin technology in enhancing asset-heavy industries.

3.3. Bringing the Metaverse closer to reality

The metaverse, a burgeoning digital landscape driven by emerging technologies, is poised to undergo a transformative leap with the integration of AI blockchain (Vaghani et al. 2022). As AI embodies procedural generation techniques, its advances promise to improve realistic environments, intelligent interactions, and content creation, fostering novel forms of creativity and making the metaverse more accessible. On the other side, blockchain solutions enable digital ownership and authentication leading to democratic governance and individuals with power over their own data. Also, they are used as backbone infrastructures promoting interoperability of ecosystems and enabling financial flows (Küpeli & Gürpinar 2023).

In the realm of virtual environments, AI is already shaping the way we interact with these digital spaces. For instance, Meta's "BuilderBot" demonstrates the potential of AI in the metaverse. This AI-powered interface allows users to conjure virtual objects into existence through simple voice commands, showcasing how intuitive and interactive the metaverse can become.

Avatars, the digital representations of users in the metaverse, are evolving to become more personalized and expressive, thanks in part to AI. Apple's Vision Pro, for instance, offers a feature enabling users to create 3D avatars for video conferencing, hinting at the diverse possibilities for self-expression in the metaverse. AI is set to expand these capabilities further, offering users a vast array of options to craft avatars that resonate more deeply with their identities and preferences.

AI's influence extends to animating the metaverse's inhabitants, from non-playable characters to virtual assistants and interactive chatbots. These AI-driven digital entities are designed to be more than mere decorative elements; they are programmed to react, interact, and tailor their responses according to individual user preferences, adding a layer of personalization to the metaverse experience.

Due to the diverse stakeholders involved in metaverse ecosystems, it is crucial to prove ownership of digital assets, including virtual goods, land, and intellectual property, fostering a vibrant economy. Here, the immutable records of blockchain solutions as well as their unique non-fungible tokens come into place that allow for transparent and tamper-proof records of transactions, interactions, and ownership changes within the metaverse, enhancing trust and accountability among users and stakeholders. Also, financial flows are transferred into the digital space and streamlined to exchange digital assets without the involvement of third parties (Huynh-The, et al. 2022).

Furthermore, on the network layer (which provides the underlying infrastructure for communication and data exchange), the decentralized infrastructures of blockchain solutions enable interoperability between different metaverse environments and platforms, allowing users to seamlessly transfer assets and identities. Ultimately, the infrastructures also enable democratic governance mechanisms enabling community-driven voting and decision-making processes regarding metaverse developments, policies, and content

creation or moderation that show potential far beyond social individual-centric metaverse approaches, but also in the industrial metaverse where enterprises co-create services and products (Küpeli & Gürpinar 2023).

A key aspect of the integration of AI and blockchain into the metaverse is their potential to enhance inclusivity (Fernandes et al.; Schuetz & Venkatesh, 2020). AI has long been instrumental in making digital experiences more accessible, such as through automated subtitles for the deaf or hard of hearing, language translation to bridge communication gaps, and providing speech assistance for individuals with conditions like ALS. With blockchain solutions, we have infrastructures at our fingertips that involve participants equally using a peer-to-peer approach and powering federated learning approaches to make data sources more diverse and relevant. In the metaverse, these capabilities are expected to be amplified, allowing for a more inclusive environment where barriers to participation are significantly reduced (Schuetz & Venkatesh, 2020).

In summary, AI and blockchain are not just changing how we perceive and interact with the digital world; they are actively redefining it. The convergence of AI and blockchain in the metaverse is set to unlock new avenues for engagement, connection, and collaboration, heralding a future where digital experiences are more immersive, personalized, and inclusive.

3.4. Sustainability & Authenticity

This bridge between AI and Blockchain holds potential for a variety of sectors and industries. Their combined capabilities can revolutionize industries by ensuring secure, transparent, and efficient operations. For example, in the international trade and supply chain industry, which impacts almost every other industry and activity worldwide, AI can predict logistical challenges, such as weather-related disruptions, while blockchain ensures transparency in product origin and handling and product authenticity. An interesting and practical example can be seen in the case of smart precision-based farming, whereby the convergence of AI and blockchain technology, can not only enhance accountability but will foster trust in global trade by offering a time-stamped ledger of transactions to the consumer. This in turn can help achieve the UN Sustainable Development Goals 2 (food for all), 3 (health for all) and 12 (sustainable production and consumption patterns) (Tyagi, 2023)

When it comes to sustainability, the fusion of blockchain and AI can significantly elevate existing solutions. For instance, in Carbon Credit Trading, blockchain verifies the authenticity of carbon credits (Espenan, 2023), while AI forecasts market trends to optimize pricing and trading (Atsalakis, 2016). In waste management, AI can anticipate waste generation, allowing blockchain to trace and incentivize source reduction (Gopalakrishnan & Ramaguru, 2019). In the realm of renewable energy, AI's forecasting abilities paired with blockchain's secure tracking can boost renewable investments (Taherdoost, 2024). Furthermore, in water conservation, AI's ability to gauge regional water requirements and wastage combines with blockchain's transparent water credit trading to champion conservation (Naqash et al., 2023). Moreover, in biodiversity conservation, AI's analytics can preempt threats in wildlife and marine reserves, with blockchain ensuring a transparent allocation of resources,

guaranteeing genuine utilization (Sivarethinamohan & Sujatha). Additionally, in smart farming, the amalgamation of AI-driven predictions regarding weather, soil quality, and crop health, paired with blockchain's guarantee of traceability, stands to revolutionize food safety and quality (Karunathilake et al., 2023). For the latest there are projects that employ IoT Automation for real-time surveillance of growth environments, dynamically activating devices to create the perfect conditions for crops. Further enhancing its capabilities, some projects use AI to process varied data, guiding farmers towards optimal decisions; the integration of computer vision offers a window into the health and quality of crops. The Digital Twin feature provides a holistic, real-time digital mirror of greenhouses by merging IoT, AI, DLT, and Virtual Reality. Additionally, through Decentralized MarketPlaces - DEX, projects can revolutionize the agricultural sector by sidestepping middleperson, thereby fostering direct links between growers and consumers and creating innovative economic avenues. An example for projects that combine these different technologies is Zignar Technologies.

AI can help sustainability by taking into proper account Environmental, Social and Governance factors, consistently measured through the blockchain. For example, ESG factors can be used to improve risk management, incentivising funds allocation towards more sustainable investments (Agosto, Cerchiello & Giudici, 2023).

3.5. Health Applications

In another example, in the healthcare space, blockchain-secured portable digital credentials can be used to store patient health records and medical history, ensuring that the patient has full control over their data while keeping the records tamper-proof. In such systems, the actual patient/medical data can be kept off-chain for privacy purposes, with cryptographic signatures or hashes stored on a blockchain to ensure transparency and tamper-proofness (Siqueira, Conceição et al, 2021). AI can analyze these records to predict patient health trends or disease outbreaks and also, if the patient chooses, the data can be donated or shared with research institutions which can use the data within an AI system to find patterns that will help them understand certain diseases or conditions. The patient, in turn, can receive an incentive (tokens) for sharing their information. For example, Halfloop provides a blockchain-secured platform where patients can store and share their medical device details, empowering clinical teams with precise data for enhanced care and offering industries real-world feedback to refine their products. In the second stage, they plan to integrate future AI capabilities. This integration of security and intelligent analysis aims not only to enhance patient outcomes and industry innovations but also to safeguard patient data privacy, all while providing incentives for data sharing.

4. Challenges and Solutions

The rapid rise of AI brings with it a series of intertwined challenges. Acquiring high-quality, unbiased data in vast quantities for AI models remains a hurdle, especially when ensuring these models generalize well in diverse real-world scenarios. The opaque decision-making

processes of many AI systems, often termed the as already mentioned the "black box" dilemma, raise ethical and trust issues, making the explainability of these models paramount (Nassar et al., 2019). Additionally, the sheer computational demands of advanced AI models present both financial and environmental concerns. As we further integrate AI into our lives, concerns about adversarial attacks compromising AI systems, the potential for job displacement, and the overarching dominance of a few tech giants in the AI arena grow more pronounced. Furthermore, the swift pace of AI advancements often outpaces regulatory measures, leading to gaps in oversight and policy. Lastly, crafting truly effective AI solutions calls for a harmonious blend of interdisciplinary expertise, a feat easier said than done.

AI and blockchain are reshaping the way businesses and society interact. Yet, Europe seems to lag in funding and development compared to global players like the United States and China (Verbeek & Lundqvist, 2021). The bulk of funding for AI and blockchain projects is concentrated in the US and China, accounting for around 80% of the total global investment of about €25 billion annually (Verbeek & Lundqvist, 2021). The European Union, on the other hand, is investing just around €1.75 billion per year, making up only 7% of the global total (Verbeek & Lundqvist, 2021). This difference reveals an annual investment gap in Europe that could be as high as €10 billion.

One key reason behind this gap is the lesser involvement of big institutional investors like pension funds and insurers, especially in supporting later-stage startups that are working on AI and blockchain technologies. This is where the crux of the issue lies - getting enough funds to help these startups move from initial concepts to fully developed, market-ready solutions. But it's not all gloomy. Europe has a strong base of high-quality research and a vast pool of digital talent. The region has the right ingredients to compete, develop, and deploy AI and blockchain technologies across various sectors. With more specialized researchers than the US and China, Europe has a strong foundation to build on.

A united effort is needed. By pooling financial resources from both public and private sectors, Europe can support the scale-up of innovative AI and blockchain ventures. This means more investments, especially in later-stage startups, to bring excellence in research to the market, helping build a smarter and greener society. With the right amount of investment and a collaborative approach, Europe has the opportunity to not only catch up but potentially lead in the AI and blockchain space, paving the way for digital transformation that could ripple through societies making them more digital, green, and sustainable. Moreover there are particular challenges of AI that can be solved or mitigated by Blockchain:

4.1. Data Privacy, Ownership, and Blockchain

AI's dependence on vast amounts of data for training and operation has raised significant privacy concerns. The General Data Protection Regulation (GDPR) sets stringent guidelines for data handling, directly impacting how AI models are developed and deployed. To elucidate the compliance with GDPR in the context of AI, we have compiled a detailed table that outlines

key stages in AI model development and their corresponding GDPR requirements (Sartor & Lagioia, 2020).

	Overview	Compliance with GDPR
Data collection	AI models typically collect data from various sources by way of web scraping, user interactions, IoT devices, or database collection. In such cases, personally identifiable information (PII) such as names, addresses and ID numbers may find their way into this data collection dragnet.	Article 5 and 6 GDPR require organizations to collect personal data lawfully and transparently with consent or other lawful bases (e.g. contract or legitimate interest) being necessary prerequisites.
Data Processing	Raw data often includes errors, missing values, or outliers. In order to clean the data, processing techniques such as imputation, outlier detection and normalization are used to prepare the data for further analysis.	Article 5 and 32 GDPR state that data processing must adhere to GDPR principles of data security, accuracy and data protection by design or default. Pseudonymisation techniques (e.g. random noise, salt and peppered hashes, scrambling) may be introduced into the query functions.
Feature engineering	Often involves aggregating or transforming data, which may inadvertently reveal sensitive information about individuals. For example, aggregating location data over time may expose an individual's routines or habits. Aggressive employment of feature engineering techniques clashes with privacy principles such as data minimization which is a prerequisite for organizations to collect and process only the minimum amount of data necessary for a specific purpose. Moreover, techniques such as Principal Component Analysis (PCA) or feature selection may inadvertently retain features containing PII. Even though this data is transformed into a lower-dimensional space, it may still contain privacy-sensitive patterns.	Feature engineering techniques must adhere to strict data minimization practices (Art. 5(1)(c) GDPR) so that only data that is strictly necessary is collected and processed. Moreover, the application of anonymisation and pseudonymisation techniques (Art. 4(5), Art.25 GDPR) should be employed to either remove all PII or otherwise replace it with pseudonyms. Privacy by design techniques (e.g. differential privacy) can be a useful technique to add noise to data, ensuring that privacy settings are configured by default. As AI models often horde data troves, it is important to establish clear data retention policies and procedures (Art. 5(1)(e) GDPR) to ensure that engineering pipelines adhere to these policies. Explicit and informed consent from individuals (Art.6 GDPR) is necessary in any and all

	Because feature engineering can be deployed for both text and image data, it can be targeted towards concepts such as named entity recognition (NER) or facial recognition/object detection. Such techniques also make use of proxy variables which can be strongly correlated with sensitive attributes meaning that personal data may not be directly used but the proxy variables can generate privacy risks.	cases where feature engineering may involve PII. Moreover, complex or high risk feature engineering may require a Data Protection Impact Assessment (DPIA) (Art.35 GDPR) to identify potential risks and propose mitigation strategies. As with any instance where data is aggregated, there is a risk of breaches. Access controls and encryption (Art.32 GDPR) need to be robust and limit who can access feature engineering pipelines and data. One of the main challenges regarding AI models and the collection of personal data is the clash with data subject rights (Art.15-22 GDPR). Feature engineering processes should in principle allow for the easy fulfillment of data subject rights – implementing mechanisms to respond to access, ratification, erasure and objection requests.
Model training	Here, datasets are divided into training, validation and test sets with different machine learning algorithms and architectures being tested to find the best-performing model. Models are fine-tuned using techniques such as grid search or random search through a process known as hyperparameter tuning. Regularization techniques like dropout or L2 regularization are also used to prevent overfitting. The data is then augmented based on need (e.g. image recognition) to increase the diversity of the training data.	Data minimization practices (Article 5(1)(c) GDPR) should be employed during model training via for example, data sampling, aggregation and anonymization. As with feature engineering it is necessary to obtain prior consent (Art.6 GDPR) and to provide individuals mechanisms to withdraw their consent. In cases where consent is not feasible, it is important to adhere to other lawful bases for data processing. The introduction of sensitive data (Art.9 GDPR) should be ring-fenced, meaning that PII such as health or biometric data requires additional protective measures. Under Article 20 GDPR, individuals may request their data be transferred over to them. As such, AI models should ensure a means to allow individuals to obtain their personal data used in model training in a structured, machine-readable format. Clear data retention policies (Art.5(1)(e) GDPR), must be in place for training data, with a means to automatically delete or anonymise the data

		once it is no longer required for model maintenance or improvement. Conducting DPIAs (Art.35 GDPR) should also be part of the process for high-risk model training activities. Last but not least, it is recommended to address and mitigate biases, ensuring fair and non-discriminatory outputs.

Blockchain technology can play a key role in addressing these GDPR compliance challenges. Its capabilities in ensuring data traceability, integrity, and secure access control align perfectly with GDPR's requirements. Moreover, as we saw previously blockchain's decentralized nature facilitates transparent and consensual data transactions, making it an ideal tool for managing data privacy in AI systems.

However, it's crucial to consider key distinctions in blockchain applications, particularly in the realm of data privacy and ownership. Firstly, the convergence of AI and blockchain technology must contemplate the type of blockchain used: Permissioned versus Permissionless. Permissioned blockchains restrict write access to a select group of entities, ensuring that only authorized participants can modify data (De Filippi & Wright, 2018). This model is often favored in regulatory and governmental applications where data sensitivity is key. In contrast, Permissionless blockchains allow any participant to write data, offering greater decentralization and openness but potentially less control over data integrity.

Secondly, the distinction between On-chain versus Off-chain data access and computation is vital from a performance standpoint. On-chain computation, where data is processed and stored directly on the blockchain, ensures higher data immutability and security. However, this can be resource-intensive and may not currently offer the ideal balance of performance and cost-efficiency for every application. Moreover, on-chain data storage can pose challenges regarding GDPR compliance due to its immutable nature, making it difficult to delete or modify personal data once it's incorporated into the blockchain. Off-chain computation involves processing data outside the blockchain while ensuring that verification processes remain on-chain, which can significantly enhance performance and scalability. However, off-chain methods might compromise data integrity and security, as the data is not fully safeguarded by blockchain's decentralized and tamper-evident structure. This could potentially increase the risk of data breaches or unauthorized access when sensitive information is handled outside the blockchain's protective ledger.

To illustrate, consider the European Union's blockchain initiatives for identity management (Meeco Group Pty Ltd, 2023)). These are typically public to read but permissioned to write, akin to how the domain name system operates where everyone can access domain data, but only registered entities can manage it. This structure allows regulated access while maintaining public transparency, ideal for applications requiring both broad accessibility and stringent control.

Furthermore, blockchain can enhance the process of obtaining and managing user consent, ensuring that data used in AI models is both legitimate and compliant with regulatory standards. This harmonization of blockchain and AI not only addresses privacy concerns but also paves the way for more ethical and responsible AI development. However, not always is consent the best solution: with generative AI, for example, the specific uses are often unknown, users of generative AI tools can use them for a myriad of different uses, some benign, some malignant. Consent becomes a "blank check" to do nearly anything, so people have no idea what they are consenting to. (Solove, 2024). Considering all the six lawful bases to gather the data about the data subject could be another possible solution to legitimate AI models. Considering all the six lawful bases outlined in Article 6 of the GDPR provides a more robust framework for data handling. These include consent, performance of a contract, legitimate interests, vital interests, legal obligations, and tasks carried out in the public interest. Leveraging these varied legal grounds can ensure that AI models are not only compliant but also legitimately engage with user data across different scenarios. This approach fosters a more comprehensive strategy to address privacy concerns, support ethical AI practices, and ensure adherence to regulatory standards.

By leveraging blockchain, individuals can gain unprecedented control over their data. This empowerment goes beyond mere access; it extends to ownership and the authority to dictate terms of data usage. Blockchain enables individuals to specify who can access their data and under what conditions, bringing a new level of autonomy and security to personal data management. This feature is crucial in an era where data is a valuable commodity, and its unauthorized use is a pervasive concern.

Incorporating blockchain technology in data management fundamentally shifts the power dynamics of data ownership and control. It empowers individuals by granting them true ownership over their data, enabling them to exercise control over who accesses their information and the conditions under which it is used. This shift is particularly significant in an era where personal data is increasingly seen as a valuable asset. We will analyze this further in the following section.

4.2. Using blockchain to secure data sharing and access control

Blockchain stands as a transformative force, addressing key challenges in data privacy, ownership, and access control (Ma et al., 2021). It involves a shift from traditional data management models, where corporations predominantly control user data, to one where individuals hold true ownership of their personal data, and this acquires higher relevance when discussing AI. In the context of AI, where data is the lifeblood of learning and decision-making processes, this shift is particularly impactful. Blockchain empowers users with unprecedented autonomy over their data, enabling them to grant or deny access under specific conditions, thus revolutionizing data sharing to be more consensual and controlled.

The concept of tokenization in blockchain further elevates this dynamic. It involves representing data or data rights as digital tokens that represent ownership and rights over digital assets. This approach paves the way for new economic models where individuals can

directly monetize their data by selling or trading these tokens. In the realm of intellectual property (IP) rights, blockchain's application extends to enhancing how creators and inventors manage and monetize their IP. By tokenizing IP rights, blockchain facilitates secure and transparent tracking of ownership and usage rights, fostering fair compensation and mitigating the risk of unauthorized use. Moreover, blockchain introduces a variety of incentive mechanisms. For example, users could receive tokens for sharing their data, which can be utilized within a specific ecosystem. Smart contracts on the blockchain can also automate transactions, transferring digital currency to users in exchange for their data usage rights. By tokenizing IP rights, blockchain ensures secure and transparent tracking of ownership and usage rights, facilitating fair compensation and reducing the risk of unauthorized use, which is vital in AI development and deployment.

Furthermore, blockchain enables the establishment of user-centric data marketplaces (Rocha et al., 2022). These marketplaces offer platforms for users to directly trade their data with companies or researchers, ensuring ethical and consensual data usage in AI applications. Users retain full control and transparency over who buys their data and for what purposes, thereby aligning AI development with ethical and user-centric practices.

The use of digital identity solutions represented by cryptographic key pairs (public and private) could be a means to overcome data-sharing challenges involving AI models. With a blockchain-based decentralized identity management system, each participant in the AI model development process (including data providers, AI developers and users) could obtain a unique digital identity on the blockchain. These identities could be pseudonymous and cryptographically secured, ensuring that only access to the private key grants access to the digital identity. For PII such as healthcare or patient data, it would be further possible to store this on a permissioned blockchain network, restricting access to such data and providing additional cybersecurity guardrails.

To generate a zk-SNARK proof for a computation on private data in an AI model, the prover would first convert the computation into a polynomial representation. This polynomial would then be used to create a ‘commitment’ to the private data without revealing it. The prover would then construct a succinct proof of knowledge for the polynomial’s consistency, which can be thought of as proving that the commitment hides a valid computation. The verifier, without knowing the private data could then check the validity of the proof and the consistency of the commitment. If the proof is valid, the verifier can be assured that the computation was performed correctly on the private data without learning anything about the data itself. SNARK proofs are also extremely compact, often just a few hundred bytes in size, regardless of the complexity of the computation (due to the use of advanced cryptographic techniques like the ‘Pinocchio’ protocol). Additionally, their non-interactive nature can simplify continuous verification in AI deployments, while their widespread adoption across various privacy technologies makes them readily available.

On the other hand, zk-STARKs are built upon more algebraic and error-correcting code-based cryptographic primitives which can provide post-quantum security and scalability. Generating a STARK proof involves encoding the computation and data into algebraic

polynomials not dissimilar to zk-SNARKs. However, STARK proofs use error-correcting codes to efficiently handle large-scale computations (Ben-Sasson et al., 2018). STARK proofs are verified by checking that they satisfy certain mathematical relations. This verification process is transparent and does not require the knowledge of the private data in question making STARK proofs both highly transparent and easily auditable. The scalability afforded by STARK proofs also makes them ideal for AI applications involving complex models and massive datasets. Unlike SNARK proofs, STARKs do not require a trusted setup, increasing their overall security profile. They are also designed to be post-quantum secure, which in a world populated by ever more sophisticated AI algorithms is imperative.

4.3. Using blockchain to improve data provenance and audit trails

In a blockchain data is stored in a series of blocks, with each block containing a cryptographic hash of the previous one. This design ensures data immutability – once data is recorded in a block, it cannot be altered without changing the subsequent blocks in the chain (this process may be different for 'blockless' protocols such as Algorand). Applying immutable ledgers to AI models could guarantee that data records and audit series are tamper-resistant, barring unauthorized attempts to alter the data or audit information without the right access controls. Blockchains record data transactions in decentralized and distributed ledgers. Each transaction includes a digital signature from the sender, details about the data and a unique transaction ID. In a blockchain, digital signatures based on public-key cryptography are used to verify the authenticity of blockchain transactions. Each participant in the network possesses a private key to sign their transactions and others can verify the signature with the corresponding public key. In AI, cryptographic signatures could ensure that data transactions and audit entries on the blockchain are verified as originating from legitimate participants thereby enhancing the audit trails integrity.

Blockchain transactions are also timestamped using the network's native consensus mechanism. This timestamp reflects the moment transactions were added to the blockchain. In an AI model context, timestamping may be essential for maintaining the chronological order of data transactions, allowing for proper data auditing and adhering to GDPR requirements such as specific data retention periods. Regarding data linkage, blockchain transactions can include references or cryptographic hashes pointing to off-chain data storage or computations allowing for data integrity while keeping sensitive data off-chain. For AI models, data linkage could ensure that the blockchain audit trail is connected to the actual data sources or model versions, enabling verification that computations were performed correctly on specific data without revealing the data itself.

In permissioned blockchain networks, auditors and regulators may be granted controlled access to the network's audit trail for independent review purposes. Auditors could leverage blockchain audit trails to assess AI model compliance with data privacy regulations and ethical standards. This dovetails with the fact that smart contracts can automate audit-related processes (e.g. altering data owners or initiating data deletion). In AI, smart contracts can automate compliance checks, ensuring data usage conforms to regulations

such as automatically enforcing data retention policies, and notifying stakeholders about data-related events.

4.4. Using blockchain to combat disinformation

At the AI For Good Global Summit in Geneva, numerous insights emerged about the profound capabilities and concerns of generative AI models (GenAI). These models, especially with milestones like GPT-4, have the potential to reshape countless sectors, bolster human creativity, and be pivotal in achieving the UN's SDGs. However, unchecked advancements in AI could amplify existing challenges surrounding disinformation. With the rapid rise of large language models, demands for AI regulation have become increasingly vocal. It's essential not to let overarching debates about the potential existential threats of AI divert attention from more immediate issues, such as the proliferation of hate speech, discrimination, and the perils of mass surveillance.

Disinformation, albeit an age-old issue, has seen its impacts amplified through social media and advanced AI technologies. Events like the 2016 U.S. elections and the Brexit referendum serve as poignant reminders of the erosion of trust in democracy and institutions (Bader, 2018). The capabilities of GenAI are ushering us into an era where distinguishing between real and "synthetic" media might become exceedingly challenging, especially when such tools are harnessed by state-sponsored entities and malicious actors, therefore more effort in critical thinking type of education must become a priority.

Navigating content moderation in the digital age is intricate. Even harmful fabrications can often be shielded under the banner of free speech. It's necessary to recognize the inherent challenges in setting universally accepted standards and mechanisms to identify and counteract disinformation, especially given the potential shortcomings of even the most advanced automated systems. A promising avenue to achieve this is the bridge with blockchain. Instead of fixating on imperfect detection mechanisms, a decentralized blockchain approach could provide a more reliable means to verify sources. Utilizing processes that do already exist in the media industry, such as forensic watermarking of content and other types of identification of original material, additional crypto hardware anchors in media production devices and the use of blockchain as a registry layer for tracking all use of AI-altered or -produced content in the news press can solve many of the issues and should become a standard (Kudelski Group, 2023).

Additionally, DAOs (Santana & Albareda, 2022)[2] and dApps (Goel et al., 2022)[3] can help to craft standards of trustworthiness that stand as transparent, immutable, and independent

[2] Decentralized autonomous organizations (DAOs) are blockchain-based entities managed by a peer-to-peer (P2P) network of contributors. They operate without central executive teams, using automated rules in smart contracts. Governance is autonomous, combining on-chain and off-chain mechanisms to facilitate community decision-making

[3] Decentralized Applications, or dapps, are web applications that operate on blockchain technology and decentralized peer-to-peer (P2P) networks, rather than relying on a single server or centralized database. This decentralized nature ensures the internet remains a public resource, accessible and controlled by many rather than a few actors.

of centralized control providing potential ranking on the level of trust that we can give to data, according to the sources, collection and processes similar to how it works in DMRV projects (Demia, 2024).

4.5. Using blockchain to improve consent management with AI models

Blockchain can be used to create a tamper-resistant and auditable record of consent. When an individual provides consent for their data to be used in AI models, this consent is recorded as a transaction on the blockchain and may include details such as the data owner's identity, the specific data elements or processing activities to which consent is granted, and the timestamp of consent. Consent transactions could be signed cryptographically by the data owner using their private key, ensuring the authenticity of the consent record. Moreover with AI, cryptographic signatures would verify that the consent transaction was indeed initiated by the data owner – preventing unauthorized use of data and ensuring that AI models are only trained or deployed with explicit, verifiable consent. The same can be achieved through the use of verifiable credentials (VCs), where consent is provided by an owner, while preserving their privacy.

Smart contracts or VCs can be further leveraged to automate the enforcement of consent rules. These contracts could contain predefined conditions for data usage based on consent, such as restrictions on the duration of consent or the specific purposes for which data can be used. If an AI model tries to access data without valid consent or attempts to exceed the scope of consent, the smart contract could block access and trigger notifications to relevant parties. This works in the reverse as well.

Blockchains could record not only the granting of consent but also its revocation. When a data owner decides to withdraw their consent, a revocation transaction could be added to the blockchain or managed through trust revocation registries used to verify VCs authenticity and validity. Consent revocation is critical in AI models to respect the data owner's wishes. Blockchains could ensure that revocation is immutable, providing a clear record of when and why consent was withdrawn. Applying blockchain to AI models can also allow for more granular consent management. Data owners could specify precisely which data elements or processing activities they consent to, enhancing control and transparency. It is also possible to link consent records to corresponding data via blockchain networks, allowing for easier data portability. In an AI context, data portability is essential to preserve the data subject's rights and blockchain could ensure that consent information remains connected to the data – simplifying data transfers while preserving consent details.

4.6. Using blockchain to improve data retention and deletion with AI

It is possible to store data retention policies as smart contracts which define how long data should be retained based on predefined criteria such as data type, sensitivity or specific legal requirements. This process can be adopted where data retention policies for AI models are encoded into smart contracts. Once these policies are recorded on the blockchain they become immutable, meaning no one including the data controller can tamper with or modify

them without leaving a clear trace. Applied to AI models, the immutability ensures that no one with access to the model can try to manipulate the data retention policies. Moreover, when data used in AI models reaches the end of its retention period or fulfills other conditions, smart contracts can be deployed to automatically initiate the deletion process. In order to bolster auditable data deletion, the transparent nature of blockchain networks makes it easier for auditors and data subjects to gauge how data is being used and processed. This becomes crucial when considering data processing in AI is often complex and can span various models and systems. Additionally, as AI models evolve, data retention policies can be adapted on the blockchain (e.g. via pointer contract) to reflect new requirements or ethical standards while maintaining a transparent history of any policy changes.

4.7. Combining blockchain and AI to improve data marketplace and incentives

One of the more forward-looking fusions of blockchain and artificial intelligence models is in the sphere of data marketplaces. A blockchain-based data marketplace could serve as a decentralized platform where data providers can securely offer their datasets, and data consumers can access and purchase them. Such a marketplace could be extended to artificial intelligence, allowing AI developers and organizations to acquire diverse datasets for training and evaluation purposes while maintaining a clear record of data transactions. By tokenizing data and transactions within the marketplace, data providers could receive tokens in exchange for their datasets. Specific data licensing and usage terms could be represented within the smart contracts, enabling clearly defined usage terms for data transactions. This would put guardrails on the means of data usage by AI developers. Blockchains can incorporate mechanisms for data validation and quality assurance, with consensus mechanisms and oracles potentially verifying the authenticity and accuracy of any data before it is listed in the marketplace.

As reliable data is critical for AI model training to prevent black boxes and data decomposition, the application of distributed ledgers to data fidelity would be very helpful in reducing the risk of training models on erroneous or malicious data sets. On-chain reputation through Soul-Bound-Tokens (SBTs), on-chain or off-chain Verifiable Credentials (VCs) or other reputation-based systems could help ensure trust in the data marketplace, allowing developers to make informed decisions about data providers based on their reputation. Data privacy could be maintained through various encryption techniques, allowing data to be encrypted both at rest and in transit. An incentive mechanism, either through token rewards as previously mentioned or otherwise royalties based on data usage and/or participation in the training process for AI models could be executed via smart contracts that automate the distribution of incentives. Smart contracts could also be used to embed the licensing terms of data that is available to AI developers. Decentralized Autonomous Organisations (DAOs) or other forms of on-chain governance mechanisms could further facilitate decision-making and dispute resolution.

However, under EU data protection law, personal data is not considered a tradable good in the traditional economic sense. Instead, personal data is treated as a special category of information (a non-rival good) with specific legal protections and restrictions as the above sections have highlighted. Therefore, the introduction of a data marketplace for B2C or C2C data would face considerable legal challenges, notwithstanding the efficiency gains borne from harmonizing these two disruptive technologies. Creating a decentralized data marketplace for B2B data could theoretically be possible. The EU Data Act introduces further requirements on smart contracts used in the application of data-sharing agreements, which a decentralised data marketplace would essentially be a constellation of. AI algorithms used in the context of data-sharing agreements under the Data Act is an area that may garner considerable attention from both technologists and the legal community as the legislation enters into force.

A workaround to this would be building data marketplaces that only allow data that has individual consent. Blockchain theoretically can allow such ecosystems where the origination of the data starts from the individual giving consent and being rewarded for that consent, going through to aggregated data that's suitably anonymised and then makes its way to a "Trusted Data Marketplace".

4.8. zk-ML: merging zero-knowledge proofs with machine learning on Ethereum

Blockchains such as Ethereum have enabled the creation of smart contracts, expanding the capabilities of code definition. However, the limitations of blockchain computation and the transparent nature of blockchain operations hinder the development of compute-heavy applications involving private or sensitive data, such as machine learning. In typical supervised machine learning scenarios, inputs are fed into a trained model, producing outputs that downstream entities utilize. With lightweight machine learning frameworks like ONNX, inference can now occur on edge devices like mobile phones or IoT devices without sending sensitive inputs to centralized servers, improving scalability and privacy.

Yet, challenges arise. There's often a need to conceal inputs and/or model parameters from public view, especially when they contain sensitive data like personal financial or biometric information. Additionally, downstream entities require assurance that the input was correctly processed by the ML model to yield the claimed output. This is where ML combined with zkSNARK protocols offers a novel solution. By utilizing zero-knowledge proofs, it becomes possible to verify computations on private data without revealing the data itself. This satisfies the contradictory demands of data privacy and computational verification, allowing for secure and private execution of machine learning tasks on blockchain platforms like Ethereum.

Consider a scenario where a consortium of healthcare institutions wants to collaboratively train a machine learning model to predict the progression of a particular disease while ensuring the privacy of patient data. Each institution holds a large dataset of patient records

including demographics, medical history, lab results, and imaging scans. However, due to privacy regulations and ethical considerations, sharing this data directly is not feasible. Each healthcare institution encrypts its patient data using homomorphic encryption techniques, allowing computations to be performed on the encrypted data without revealing the underlying information. They then use zk-ML techniques to collaboratively train a machine-learning model on the encrypted data. zk-ML ensures that the training process remains private and that sensitive patient information is not exposed. Instead of sharing the entire encrypted dataset, each institution generates succinct proofs, such as zk-rollups, to provide a cryptographic summary of their data. These proofs may contain aggregate statistics including averages, variances or gradients derived from the encrypted patient records. By utilizing zk-rollups, the computational load is significantly reduced compared to traditional methods as only the compact proofs need to be transmitted and verified.

Consider another example where financial institutions need to detect fraudulent transactions while preserving the privacy of sensitive customer data. The first step could involve encrypting customer transaction data using homomorphic encryption libraries like SEAL or HElib, ensuring the data remains encrypted throughout the training and inference process. Following this, machine learning algorithms supporting encrypted computation could be used (e.g., Microsoft SEAL's encrypted neural network library or PySyft's federated learning framework with encrypted aggregation) to train the model on the encrypted transaction data and discern patterns indicative of fraudulent behavior. Zero-knowledge libraries such as libsnark or ZoKrates could be used to generate proofs demonstrating the accuracy of the model's predictions without revealing sensitive data or model parameters. Deployment could involve the integration of the zk-ML model into financial institutions' fraud detection systems where it can analyze encrypted transaction data in real-time. When a potentially fraudulent transaction is detected the system could then generate alerts or take action while ensuring customer privacy is maintained.

4.9. Extending neural networks to zk-proofs

A neural network is a computational model inspired by the structure and functioning of the human brain's interconnected neurons. It consists of a network of artificial neurons, also known as nodes or units, organized into layers. Each neuron receives input signals, processes them using a set of weighted connections, and produces an output signal. Neural networks are used for various tasks including pattern recognition, classification, regression and sequence generation among others.

A neuron (node/unit) is the basic computational unit within a neural network. It receives input signals from other neurons or external sources, computes a weighted sum of these inputs, applies an activation function to the sum and produces an output signal. Connection edges link between neurons through which signals propagate. Each connection is associated with a weight - determining the strength of influence of the input signal on the neuron's output. Neurons within a neural network are organized into layers. There are typically three types of layers:

- Input Layer: Receives input signals from the external environment or previous layers.
- Hidden Layers: Intermediate layers between the input and output layers. They perform complex computations by transforming input signals into meaningful representations.
- Output Layer: Produces the final output signals of the neural network.

Weights and biases are parameters associated with connections and neurons, respectively. Weights determine the strength of influence of input signals on neuron outputs, while biases provide an additional adjustable parameter that helps control the neuron's activation threshold. Activation functions are nonlinear functions applied to the weighted sum of inputs to introduce nonlinearity and enable the neural network to learn complex patterns and relationships. Common activation functions include sigmoid, tanh, ReLU (Rectified Linear Unit), and softmax. Feedforward propagation concerns the process of propagating input signals through the neural network from the input layer to the output layer – layer by layer - without feedback loops. It computes the output of each neuron and passes it as input to the neurons in the subsequent layer. Backpropagation on the other hand is an optimization algorithm used to train neural networks by adjusting the weights and biases based on the difference between the predicted outputs and the actual outputs. It involves computing gradients of the loss function with respect to network parameters and updating them in the opposite direction of the gradient.

Extending the benefits of neural networks, particularly their multi-layered architecture, within zkSNARKs poses several challenges from technical, computational, and cryptographic perspectives. zkSNARKs are designed to prove the correctness of computations succinctly, but they impose strict limitations on computational complexity and memory usage. Implementing multi-layer neural networks within zkSNARKs requires optimizing computations to fit within these constraints while maintaining the desired level of accuracy. Additionally, neural networks often involve complex mathematical operations such as matrix multiplications and nonlinear activation functions. Performing these operations within zkSNARKs using homomorphic encryption techniques incurs significant computational overhead, making it challenging to achieve efficient execution.

Designing arithmetic circuits for zkSNARKs that accurately represent the computations of multi-layer neural networks while remaining tractable is non-trivial. The circuit complexity increases exponentially with the number of layers and neurons, necessitating careful optimization and abstraction. Generating zero-knowledge proofs for multi-layer neural networks involves proving the correctness of each layer's computation while hiding sensitive data and model parameters. Achieving this efficiently and securely requires advanced cryptographic techniques and optimizations.

4.9.1. Implementing a layer 2 neural network within a SNARK proof

In order to combine these technical constraints it may be necessary to implement a layer 2 (Gangwal et al., 2023)[4] neural network fully inside a SNARK proof. This would require designing arithmetic circuits that represent the computations of each layer of the neural network (including matrix multiplications, bias additions and activation functions). Efficient data structures and algorithms would be required to minimize circuit complexity while accurately capturing the neural network's behavior. Fully homomorphic encryption (FHE) or partially homomorphic encryption (PHE) could be used to perform computations on encrypted data within the SNARK. It is important to choose encryption parameters and algorithms that balance security and efficiency for the given application.

rust Copy code

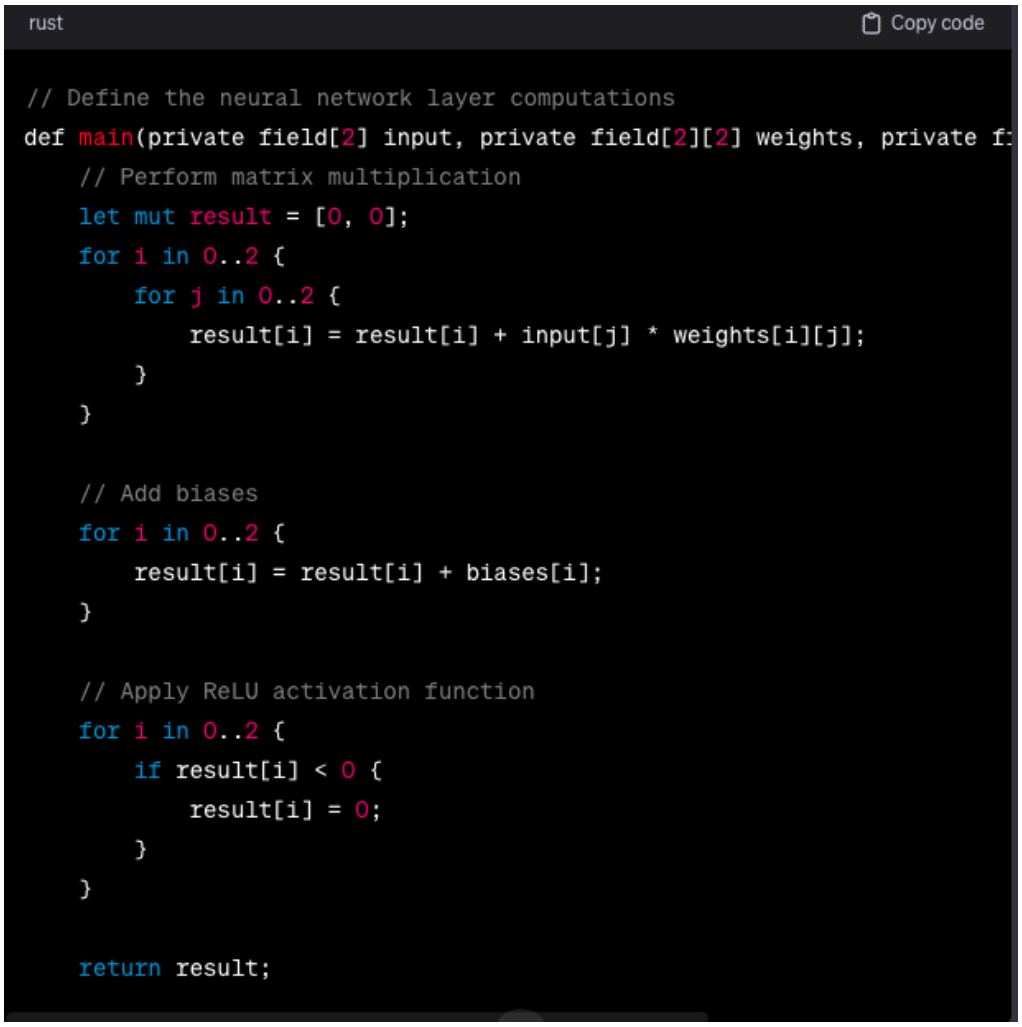
```
def main(private field a, private field b, private field c) -> (field):
    # Constraint for matrix multiplication
    let mul_result = a * b

    # Constraint for bias addition
    let add_result = mul_result + c

    # Constraint for ReLU activation function
    let output = if add_result > 0 { add_result } else { 0 }

    return output
```

Defining circuit constraints

Optimization techniques such as batching, parallelization, and algorithmic optimizations could help reduce computational overhead and memory usage. This includes optimizing matrix operations, activation functions, and other neural network primitives for efficient execution within SNARKs. It is also necessary to develop specific algorithms and protocols for generating zero-knowledge proofs that demonstrate the correctness of neural network computations while preserving privacy and confidentiality. Here, SNARK-friendly techniques such as polynomial commitments, succinct argument systems, and efficient proof aggregation could be employed to minimize proof generation time and size. Finally, it is

[4] Layer 0, Layer 1, and Layer 2 in blockchain architecture represent different layers of functionality and scalability enhancements. Layer 0 includes the foundational infrastructure, such as hardware and network nodes that support information exchange. Layer 1 is the base blockchain protocol responsible for core operations like consensus mechanisms, block validation, and transaction processing. Enhancements at Layer 1, such as increasing block size and sharding, aim to improve scalability by modifying the fundamental protocol. Layer 2 solutions, built on top of Layer 1, address scalability without altering the underlying protocol. These solutions, including payment channels, sidechains, and rollups, enable off-chain transactions to reduce the load on the main chain and enhance transaction throughput while maintaining security and backward compatibility.

necessary to implement verification procedures to validate the integrity and correctness of zero-knowledge proofs generated by the SNARK-based neural network. This will ensure that proofs are efficiently verifiable by third parties and resistant to attacks such as forgery or tampering.



Implementing a simple neural layer with ZoKrates

zoKrates DSL is used to define the computations of a neural network layer, including the matrix multiplication, bias addition and ReLU activation function. The generated arithmetic constraints can then be compiled and used to generate zero-knowledge proofs – demonstrating the correctness of the neural network layer's computations within a SNARK.

4.10. Proof of personhood and non-personhood

AI has reached a level of sophistication where it has started to simulate human behavior, and that too convincingly. This makes distinguishing between AI bots and genuine human interactions difficult, and as time continues, the problem will continue to get worse as AI

gets smarter. This can pose significant risks, intentional and unintentional, across various domains like digital finance, social media, customer service, and content creation.

The intentional risks are posed by bad actors that want to use AI's ability to simulate human behavior for fraud, scams, misinformation, and market manipulation. Risks Posed by Human-Mimicking AI

- Identity Theft and Fraud: Through the use of AI or AI bots, one could impersonate real individuals to commit fraud, access restricted services, or manipulate online systems designed for human users. Example impersonating individuals for phishing scams or in web3, claiming airdrops designed for genuine human participants.

- Market Manipulation: Beyond cryptocurrency, AI can influence stock or commodity markets through coordinated trades or spreading misleading information. Example influencing stock markets through artificial hype.

- Misinformation and Propaganda: Automated accounts spreading fake news can sway public opinion, disrupt elections, or incite social unrest on a massive scale.

- Trust Erosion: The difficulty in distinguishing AI from human interactions may undermine trust in digital communications, affecting online commerce, information reliability, and social engagement. Example AI bots might be used to create and promote fake product reviews, or online content which may or may not be accurate.

Blockchain-based identity verification systems e.g. Verifiable Credentials (VCs) can help solve this issue through providing proof of personhood.

VCs are a digital, cryptographically secured version of a credential and hence fraud-proof, within reason and depending on purpose. VCs are issued by an organization (or likely an individual) to a holder for a particular purpose. The holder of these digital credentials have full ownership and control of the VC as well as the data within it; hence they are privacy preserving.

The purpose of VCs, for example, can be:

- An identity certificate
- Holding a particular title or role in an organization
- Having a particular qualification, skillset or expertise
- Having attended a conference or seminar

The holder can share an entire VC or only selected data from it, offering zero knowledge proofs, e.g. a user can prove their age or membership in a group without disclosing their birth date or identity details.

VCs can in turn be used to issue Human-Only Credentials, providing proof of personhood. However trusted entities or community consensus mechanisms would need to take on this responsibility. These could be based on in-person verifications, biometric data checks, or similar other methods that are challenging for AI to mimic.

This could mean easily gaining access to platforms, goods or services based on VCs and proof of personhood, reducing the risk of bots exploiting systems designed for humans.

By leveraging blockchain for decentralized identity verification and utilizing verifiable credentials, digital ecosystems can significantly enhance their ability to differentiate between AI bots and genuine human users. This approach not only protects against fraudulent activities and ensures the integrity of online interactions but also upholds privacy and data sovereignty for individuals in the digital space.

The rapid progress in AI has also led to AI agents and bots that are increasingly taking on roles traditionally performed by humans, on behalf of humans. We are already seeing AI agents (such as AutoGPT) tasked with finding jobs, completing taxes, or planning trips. Very soon, these AI agents will be able to undertake more complex operations making daily tasks a breeze. The AI agents may need to interact not only with humans but also with AI and other AI agents.

Bad actors would necessarily want to take advantage through, e.g. rogue AI agents or bots. Such malicious attempts would need to be restricted by gating such rogue agents so that only legitimate AI agents are allowed access.

Hence, the need for proof of non-personhood will emerge, in a similar fashion to that of proof of personhood, required to identify and authorize only legitimate AI agents that perform tasks for their human counterparts and have the authority to do so; similar to a parent providing a letter for their child to attend a school trip, individuals could issue digital credentials to their AI agents. These credentials would not assert personhood but rather confirm the agent's authorized status to act on behalf of the user.

For instance, when faced with online services that block non-human users to prevent bot-driven abuse, such as a travel booking website, a verifiable credential could serve as a passport for AI agents. This credential would effectively communicate that while the agent is not a person, it is operating under the explicit authorization of a legitimate account holder. This flips the traditional verification paradigm, acknowledging the legitimacy of AI actions in a controlled and recognized manner.

As we navigate the complexities of a digital age where AI agents increasingly act on our behalf, the development of proof of non-personhood through verifiable credentials emerges as a vital solution. This approach not only facilitates the seamless operation of AI agents

within human-centric platforms but also ensures a balance between innovation and integrity in digital interactions.

5. Ethical Social and Governance (ESG) Considerations

Ethical and regulatory concerns individuals and then we can talk about the convergence. Recent regulatory documents and discussions, such as the EU AI Act, have introduced the need to develop AI applications that are responsible, that is, that can measure the risks they generate and, therefore, manage them, in line with the stakeholders' risk appetites. In other words, organizations that provide or develop AI systems should accompany them with an appropriate AI risk management model, that can help organizations to measure, manage and mitigate risks, throughout the whole AI lifecycle, from use case design to continuous monitoring in production. For more details see, for example, the paper by Giudici & Raffinetti (2022)., who propose a risk management system built upon four S.A.F.E. principles: Sustainability, Accuracy, Fairness and Explainability. For each principle, they propose metrics that can be employed in practical use cases for AI risk management. As mentioned in Section 4, blockchain technology offers significant potential benefits for AI. It can increase the trustworthiness of AI decisions, improve the quality of data used in AI systems, expand the capabilities of AI, and create additional value by enabling individuals to contribute their data to model training as part of a data economics model. Implementing these mechanisms at a societal level raises ethical and practical concerns that go beyond considerations like public readiness, educational levels, and industrial preparedness. Most notably, it involves assessing the readiness of law enforcement across political boundaries to regulate and prevent potential issues arising from the convergence of the emergence of AI and Blockchain.

Blockchain and digital ledger technologies enable programs to function as independent non-human entities that interact with human individuals and organizational entities. These technologies, being automated and devoid of consciousness, operate according to predefined rules and cannot adapt to unforeseen circumstances.

5.1. Governance and Voting Assistance:

AI can assist by providing educational resources and information to voters. Chatbots and AI-driven interfaces can answer questions about proposals, voting mechanisms and governance protocols. AI could be leveraged to assess the potential impact of governance proposals by considering historical data and community feedback, allowing it to estimate the economic and technical implications of the proposal. By aggregating and summarizing community feedback on governance proposals from various sources (e.g. social media, forums and chats), AI can help provide voters with a concise overview of the arguments

for/against proposals, making it easier to assess community sentiment. Additionally, AI-powered dashboards can track the progress of governance proposals, including the number of votes cast, quorum reached and voting trends.

Prediction markets are another opportunity where DeFi can leverage artificial intelligence. Prediction markets allow users to speculate on the outcomes of governance proposals, allowing voters to gauge community expectations and make more informed voting decisions. AI-driven smart contracts can be further used to automate the execution of governance proposals, reducing administrative burden on token holders. One of the more important implications of AI on DeFi is the means to ensure voter identity verification to prevent fraudulent votes while maintaining user privacy. In such cases, AI can be used to ensure only eligible token holders participate in governance votes.

5.2. Trust and Adoption

Both artificial intelligence and blockchain technology rely on the social dimension of trust in order to increase adoption via network effects. The more a technological substrate is trusted by potential users, the greater the likelihood that it will scale concomitantly. By the same token, if a technology proves to be insufficient in meeting the needs and expectations of users, it faces hurdles in its widespread adoption. If one applies the social contract theory of Jean-Jacques Rousseau, it posits that individuals willingly enter into a social contract, surrendering certain aspects of their natural freedom in exchange for collective benefits and security derived from a just government. Applying this perspective to the amalgam of AI and blockchain technology, the decentralized and transparent ethos of blockchain becomes a digitized reflection of the social contract. Participants engaging in blockchain networks effectively form a digital societal structure where they collectively agree to operate within a system characterized by transparency, immutability and decentralization. In a parallel to Rousseau's social contract – these participants entrust their data and transactions to the blockchain, placing reliance on the system's inherent integrity and transparency to foster a sense of collective trust. Within this context, the blockchain serves as a digital governance structure, where the immutability and transparency of recorded transactions function as the foundational pillars of trust. The distributed ledger (embodying the principles of Rousseau's just governance) offers participants a verifiable and auditable record of all interactions within the network. In effect, a web of trust through blockchain becomes the bedrock for artificial intelligence. The introduction of AI algorithms underscores the need for the preservation of ethical standards. Conversely, the transparency inherent in blockchain serves as a safeguard – enabling stakeholders to audit and assess the ethical implications of AI decisions, ensuring they align with a collectively agreed-upon set of ethical standards.

From an economic perspective, the decentralized nature of blockchain networks and the programmable logic of smart contracts set the stage for participants to engage in strategic interactions with predefined rules – following the game theoretic logic of John Nash. This dynamic equilibrium mirrors the foundational ideas from Friedrich Hayek's market process

theory, where blockchain facilitates secure, transparent and automated transactions – aligning towards an efficient market mechanism. Moreover, integrating AI with blockchain introduces optimization mechanisms aligned with the bounded rationality of Herbert Simon. AI's ability to optimize decision-making processes and automate intricate tasks enhances economic efficiency within a decentralized setting. As participants engage in self-interested actions within the blockchain's predefined rules, the resulting synergy also exemplifies principles of market-oriented economies – resonating with the theoretical underpinnings of scholars like Milton Friedman.

Taken together, merging the social and economic benefits of artificial intelligence and blockchain hinges on the establishment of transparent, ethical and accountable frameworks. Transparency ensures that the processes and decision-making algorithms used in AI models are understandable and auditable. Ethical considerations must uphold both the normative and legal axioms that govern our body politic, while accountability mechanisms must be in place to address unintended consequences and ensure responsible behavior.

5.3. Education and Literacy

The integration of AI and blockchain into educational programs at all levels, from schools to universities and executive education courses, is becoming increasingly essential in today's rapidly evolving digital landscape. As technologies continue to reshape societal structures and industries, there arises a critical need for individuals to develop a comprehensive understanding of their capabilities, challenges, implications, and ethical considerations. Therefore, new educational courses tailored to AI and blockchain are essential to equip learners with the necessary skills and knowledge to navigate and thrive in the digital age (Fomin et al. 2024).

At the core of these educational initiatives lies the recognition of the interdisciplinary nature of AI and blockchain (Düdder et al. 2019). These technologies intersect with various fields, including computer science, information systems, economics, supply chain management, law, ethics, and sociology, among others. Hence, fostering an interdisciplinary understanding and cooperation becomes imperative. This is reinforced by the fact that various organizational functions and departments have to be involved in industry implementation projects. Only by integrating diverse perspectives and expertise, individuals can better comprehend the multifaceted challenges and leverage opportunities presented by AI and blockchain, as well as devise holistic solutions that address complex real-world problems effectively (Gürpinar et al. 2024; Düdder et al. 2021).

In educational settings, interdisciplinary courses that combine AI and blockchain curricula can provide students with a comprehensive understanding of both technologies' technical aspects, applications, and societal impacts (Düdder et al. 2021). Moreover, emphasizing interdisciplinary cooperation cultivates collaborative problem-solving skills, encouraging learners to work across disciplinary boundaries to tackle multifaceted challenges effectively. Such collaborative experiences not only enrich learning experiences but also mirror

real-world scenarios where professionals from diverse backgrounds must collaborate to innovate and address complex issues.

Furthermore, instilling an ethical framework within AI and blockchain education is crucial. As these technologies exert significant influence on society, learners must develop a deep appreciation for ethical considerations, such as data privacy, algorithmic bias, and transparency. By integrating ethical discussions and case studies into educational curricula, students and professionals need to learn to approach AI and blockchain development and deployment with mindfulness and responsibility, thus contributing to the creation of a more ethical and equitable digital future.

5.4. Acceptable economic models

What makes an economic model acceptable and what matters more, ethical considerations or efficiency gains? In dealing with AI-driven economic models it would follow that both in equal measure necessitate careful consideration. On the one hand, an acceptable economic model should align with ethical frameworks that prioritize fairness, transparency and the well-being of individuals within society. Philosophers like John Rawls (1958), with his concept of 'justice as fairness,' emphasize the importance of minimizing social and economic inequalities to create a just and equitable society. Applying this maxim to AI-driven economic models entails ensuring that the benefits and risks are distributed fairly among all stakeholders. For example, AI algorithms in hiring processes can be designed to mitigate biases, ensuring that opportunities are accessible to diverse candidates based on merit rather than perpetuating systemic inequalities. Yet AI models trained on datasets may perpetuate and even exacerbate existing social biases – leading to discriminatory outcomes such as in the criminal justice system. In healthcare, AI-powered diagnostic tools, when designed with ethical considerations, can enhance medical accuracy, speed up diagnoses, and improve treatment outcomes, thereby upholding individuals' right to quality healthcare. In another vertical, the use of AI in surveillance systems may exacerbate privacy challenges. Facial recognition technologies employed by governments may infringe on individuals' right to privacy, prompting ethical debates around the balance between security and personal freedoms.

Economically speaking, AI has the potential to unleash a new frontier of human flourishing, by taking over remedial and monotonous tasks, freeing up precious time for human beings to engage in more creative and self-fulfilling pursuits. It could also lead to dystopian outcomes by rendering humans obsolete while the social safety nets designed to absorb less productive human economic output struggle to keep pace with the employment displacement. This could exacerbate social tensions and put pressure on governments' social and fiscal policies. Taking a neoclassical economic approach, an economical model should leverage AI to optimize decision-making processes, enhance productivity and contribute to overall economic growth. The model should also align with principles of innovation, encouraging the development of cutting-edge technologies to foster economic progress. For example, using AI-driven predictive analytics to minimize waste, reduce costs

and enhance operational efficiency in industrial supply chains, or using AI-powered robots to develop new products, improve production processes and increase productivity in the manufacturing sector. Left unchecked, this may also lead to concentrations of economic power, particularly by corporations with substantial resources. These entities can leverage AI to enhance market dominance, creating barriers to entry for smaller competitors. This concentration raises concerns about fair competition, limiting innovation and potentially stifling market diversity. In the tech industry, major players deploying advanced AI algorithms for services like search engines, social media platforms and e-commerce have amassed significant control over user data and market access. This dominance allows them to influence user behaviors and preferences, creating challenges for smaller companies to compete on an equal footing.

This is also because the integration of AI into business operations often involves substantial initial costs. Companies need to invest in advanced hardware, software and skilled personnel to develop and implement AI solutions. For smaller enterprises, these upfront expenses can act as a deterrent thereby hindering their ability to adopt AI and remain competitive. Building on the manufacturing example above, SMEs in manufacturing face challenges in implementing AI-powered automation due to the high costs associated with retrofitting existing systems or acquiring new technology. This initial financial barrier can impede their ability to leverage AI for efficiency gains and innovation, potentially leading to a technological divide in the business landscape. It is also the case that many incumbents especially in the Google, Amazon, Facebook, Microsoft ‘GAFAM’ category have been working on AI tooling for years. The economic moat built around their business models, coupled with their considerable market share almost guarantees their continued dominance with the rollout of AI in their product suite.

- Using blockchain to curb perverse instantiation with AI in Proof of Stake (PoS) Consensus

In a PoS blockchain system, validators propose and vote on new blocks and form a consensus on (come to a decision about) canonical blocks - these are the blocks that the network considers "valid’ - based on the amount of crypto assets they hold and are willing to ‘stake’ as collateral. In the context of AI governance, validators may need to stake additional tokens as insurance against potential AI misbehavior. This incentivizes validators to diligently monitor AI activities and intervene if necessary to prevent harmful actions. Slashing penalties (risk of losing staked crypto-assets) act as an additional deterrent against the negligence of collusion among validators regarding AI governance. Moreover, smart contracts can define the rules and parameters governing AI behavior, including permissible actions, constraints, and thresholds for intervention. These contracts may codify ethical guidelines, legal requirements, and safety protocols that AI agents must adhere to. Smart contracts can further integrate with external monitoring systems and oracles to continuously assess AI performance and detect deviations from established norms. If an AI

agent exhibits signs of perverse instantiation or harmful behavior, predefined triggers within the smart contract can initiate corrective measures or alert human operators.

Smart contracts can facilitate decentralized decision-making processes for AI governance, allowing stakeholders to collectively determine policies, update parameters, and resolve disputes. This ensures transparency, accountability, and consensus among participants in the blockchain network. They can also escrow funds or tokens as incentives for AI agents to comply with specified objectives or penalties for non-compliance. In cases of dispute or disagreement, smart contracts can trigger arbitration mechanisms to resolve conflicts fairly and efficiently.

- AI rulemaking and economic governance

In a recent article, Ethereum co-founder Vitalik Buterin outlined his optimism and concerns around the implementation of AI-based rule-making and economic governance within blockchain-based systems (Vitalik, 2024). He argues that combining crypto and AI can lead to the creation of a singleton": a single decentralized trusted AI that some application would rely on for some purpose. This has promise for improving AI safety in a way that avoids the centralization risks associated with more mainstream approaches to that problem. However, there are also many ways in which the underlying assumptions could fail, so it is important to tread carefully when deploying these applications in high-value and high-risk contexts. He also argues that cryptographic and blockchain-based techniques can be used to incentivize making better AI without completely encrypting it. For example, prediction markets can be used to incentivize AIs to make good predictions on a wide range of questions. AIs are willing to work for less than $1 per hour, and have the knowledge of an encyclopedia - and if that's not enough, they can even be integrated with real-time web search capability. By making prediction markets work on a microscopic scale, it is possible to reuse the "prediction market" primitive for many other kinds of questions.

Yet he also underscores the fundamental issue of scaling AI-based decision making, for example within the context of DAOs. The first is cryptographic overhead. Buterin argues that cryptographic tools, particularly versatile ones like ZK-SNARKs and MPC come with substantial computational overhead. While an Ethereum block can be verified by a client in a few hundred milliseconds, generating a SNARK to validate the correctness of such a block can take hours. The overhead of other cryptographic gadgets, such as MPC can be even more pronounced. He further explains that AI computation is already resource-intensive with the state-of-the-art language models (LLMs) only generating words slightly faster than human reading speed. The associated costs of training these models are often in the multi-million dollar range. In this regard, the significant disparity in quality between high-end models and those attempting to economize training costs raises legitimate concerns about the feasibility and efficacy of enhancing AI with cryptographic guarantees.

The second concern is what is known as adversarial machine learning which involves manipulating AI models by introducing carefully crafted inputs to deceive or mislead the system. This can be done by adding small perturbations to the input data that are imperceptible to humans but can cause the AI model to make incorrect predictions or decisions. Buterin argues that if an AI model used as part of the game rules is closed (meaning its inner workings are not transparent or verifiable), it can be vulnerable to adversarial attacks. On the other hand, if the AI model is open, attackers could potentially exploit its vulnerabilities by simulating and optimizing attacks offline before deploying them in the live network. An example includes when a user has access to an AI assistant within an open-source wallet. In this scenario, the inherent risk is that malicious actors could exploit the same AI assistant presenting a potential avenue to refine and optimize scams without triggering the wallet's defenses. Given that all modern AI has vulnerabilities, a training process even with limited access to the model can uncover these weaknesses. Contrasting this with the concept of 'AIs participating in on-chain micro-markets' and extending the prediction market analogy above, the approach proves more resilient. In this framework, numerous AIs operate within an open ecosystem each susceptible to similar risks but intentionally fostering a collaborative environment. The security of the system relies on the openness of the rules of the game rather than the internal workings of individual AIs, enhancing overall robustness against adversarial threats.

5.5. Global Collaboration

As with the myriad disruptions facing the world today, effectively managing the deployment and optimization of AI and blockchain-based business models cannot be effectuated in a political vacuum. It will require coordination and cooperation on a planetary scale, involving multilateral institutions and fora, and governments and the national, regional and local level. It will also require buy-in from the general population whose questions and concerns need to be addressed in a way that doesn't undermine the fundamental rights of human beings including the right to work, be free from unjust persecution and move freely without threats to their privacy. Recent months have seen significant strides in shaping the regulatory landscape for artificial intelligence on global and regional scales. Two landmark examples of these regulatory efforts are the UN AI Resolution and the EU AI Act, each setting comprehensive standards and guidelines for the responsible development and application of AI technologies.

The UN AI Resolution sets out guidelines for the development and use of artificial intelligence, aligning it with human rights and international laws. The resolution emphasizes the importance of creating AI technologies that are safe, reliable, and beneficial, particularly in promoting social and sustainable impacts globally. It also highlights the need to bridge the technological divide between developed and developing countries and supports the achievement of the UN SDGs. The resolution advocates for AI to be a tool for good, contributing positively to societal challenges while being developed and used responsibly.

The EU AI Act introduces a regulatory framework for artificial intelligence, categorizing AI systems based on their risk level. The act specifies that high-risk AI applications, such as those used for personal scoring systems, are strictly regulated or outright banned to prevent potential misuse. This approach is part of the EU's broader effort to ensure that AI technologies are developed and deployed in a manner that is safe, ethical, and respects fundamental rights, aiming to mitigate risks associated with AI and enhance user trust and safety.

Moreover, the international Telecommunication Union (ITU) – a specialized UN agency – has been actively engaged in discussions related to AI, especially in the context of information and communication technologies. The United Nations Educational, Scientific and Cultural Organization (UNESCO) has also shown interest in the ethical implications of AI, particularly concerning education and culture. Moreover the UN General Assembly and UN Economic and Social Council (ECOSOC) periodically discuss and address issues related to AI, emerging technologies and their impact on society. Global central banks, including the Bank for International Settlements (BIS) are further actively exploring the impact of AI on financial markets and cyber resilience. National governments, including the United States and China are actively exploring their own AI strategies designed to elevate the technology to a national economic and security imperative.

Furthermore, in February 2024, the Internet Governance Forum (IGF) of the United Nations spotlighted Artificial Intelligence in its Programme Development, demonstrating a significant focus on AI during the IGF 2024 Call for Thematic Inputs. This call engaged over 310 stakeholders in a comprehensive discussion during eight weeks, using an online submission system that allowed participants to choose up to three themes and associated issues from a list of eleven predefined categories. This inclusive initiative garnered 824 selections for themes and 1,549 for specific issues, highlighting the global community's engagement and interest in shaping the future of internet governance with respect to emerging technologies. AI emerged as a notable focus, capturing 17% of the mentions within the selected themes. Discussions on AI centered around several critical areas: AI ethics led with 30%, followed by AI governance at 24%, AI risks at 16%, AI and data at 14%, and AI applications and solutions at 9%. The smallest focus was on AI design and development, which accounted for 7% of the discussions. This outcome underlines the pressing need for ethical frameworks, governance standards, and risk management strategies in the development and implementation of AI technologies, aligning with global efforts to harness AI's potential responsibly.

To bring about a more coordinated global approach, it would be pertinent to consider the establishment of a specialized task force under the UN to focus on the ethical implications of AI and blockchain convergence, promoting collaboration among member states and stakeholders. This could feed into a wider Global Digital Cooperation Framework that guidelines for responsible AI and blockchain deployment across borders. At G20 level, it would be beneficial to further establish a permanent working group focused on emerging

technologies, including AI and blockchain to foster collaboration, share best practices and harmonize policies. This could be paired with a G20 Initiative for Ethical AI – aiming to create a unified ethical framework that member states can adopt in their national policies.

International standards organizations also have a role to play. Dedicated committees within the International Standards Organisation (ISO) and International Electrotechnical Commission (IEC) dealing with frameworks addressing interoperability, security, and ethical considerations in AI and blockchain technologies would support multilateral initiatives. The Internet Governance Forum (IGF) could organize thematic tracks and workshops specifically focused on AI and blockchain, encouraging open discussions and knowledge-sharing. The IGF could also establish a working group to create a set of best practices for the responsible use of AI and blockchain in the digital era. National governments should collaborate to share best practices and lessons learned in the development of policies related to AI and blockchain while encouraging the formation of national AI and blockchain advisory boards comprising experts from various domains. Regional bodies such as the European Union are already moving ahead with cross-border initiatives such as the landmark AI Act to harmonize regulations and provide an environment addressing ethical considerations while promoting innovation. Europe is already moving forward with different sandboxes for blockchain technology such as the European Blockchain Sandbox, while prioritizing funding for emerging technology including blockchain and AI through equity and non-equity based financing through the European Innovation Council (EIC).

Additionally, there are national security implications to consider, especially when discussing the impact of AI on global governance. The information age can potentially create tension between great powers where control of the resources (in this case rare earth metals) needed for chip manufacturing and technical know-how to design super-sophisticated AI systems risks a myopic race to the bottom ending in a zero-sum game. In the context of AI development, governments increasingly operate within a complex game theoretic landscape where the sharing of information is a strategic decision. The underlying dynamics are influenced by factors such as competition for technological advantage, national security concerns, and economic interests. Governments view AI capabilities as a strategic asset leading to a competitive dynamic where withholding information becomes rational. Sharing key insights or breakthroughs might erode a nation's competitive advantage in the global AI landscape. Following this, governments often perceive AI advancements as integral to national security. Disclosing detailed information might expose vulnerabilities or reveal the extent of a nation's capabilities, potentially compromising security. Additionally, the economic benefits associated with AI innovation contribute to the incentive for governments to prioritize their domestic industries. Limiting information sharing helps maintain a favorable economic position.

- Nash equilibrium challenges

Globally, achieving a Nash equilibrium (where no single government has an incentive to unilaterally deviate from its strategy) is challenging due to a lack of trust. Governments are skeptical that others will reciprocate information sharing. Governments may have asymmetric information about the true capabilities and intentions of other nations in the AI race. The rapidly evolving nature of AI technologies also introduces uncertainty and governments may be hesitant to commit to long-term cooperation when the strategic landscape can change quickly, disrupting equilibrium. In a zero-sum situation where one nation's gain is perceived as another's loss, the lack of global collaboration can lead to suboptimal outcomes for humanity. Governments, acting in isolation may duplicate research and development efforts, resulting in inefficient allocation of resources. This duplication can slow down overall progress and limit the potential benefits of shared knowledge. Without international cooperation, AI development may follow divergent paths resulting in fragmented solutions and standards. Interoperability challenges and lack of harmonization could hinder the seamless integration of AI technologies on a global scale. One of the more alarming outcomes is the lack of ethical concerns when a collaborative approach is absent at an international level. Governments pursuing the development of AI technology for divergent purposes in isolation from each other may lead to the development of AI systems with varying ethical standards and principles. In an extreme case it could lead to the development of AI systems with little or no ethical safeguards, posing an existential threat to the planet.

It is therefore paramount that the discussion and application of AI and blockchain take shape at an international level, and with a waterfall effect towards all national and regional stakeholders, private industry and global citizens. The ramifications of operating in a black box environment are big enough that it merits an unwavering commitment to transparency, fairness, collaboration and a human-centric approach to the development and implementation of these technologies. A firm commitment to such an approach is the first step to avoiding a zero-sum outcome.

6. Conclusion: The Road Ahead: Future Prospects

The convergence of AI and blockchain represents a pivotal shift towards creating more ethical, transparent, and effective technological solutions. By leveraging the unique strengths of each technology, we can address some of the most pressing challenges faced by digital innovations today, including concerns around privacy, security, and ethical decision-making.

AI brings unparalleled capabilities in data processing, pattern recognition, and predictive analytics - driving efficiency and innovation across various sectors. However, its challenges, such as the potential for bias, lack of transparency, and privacy issues underscore the need for a complementary solution. Blockchain, with its decentralized nature, immutability, and secure data management, offers the perfect counterbalance to AI's limitations. It ensures transparency, enhances data security and empowers users with control over their information.

Together, AI and blockchain are not just transforming industries but are also setting new standards for ethical technology development. Their integration fosters a more accountable and trustworthy digital ecosystem, where decisions are not only data-driven but also socially responsible and aligned with ethical standards.

Furthermore, a historical parallel can be very useful in understanding the potential of this combination. Also in the history of the Internet, it was the convergence between this and various technologies, such as smartphones, social networks and big data, that bequeathed the great social transformations we are experiencing today. Swap "Internet" for "blockchain" and we have a good chance of seeing a similar development in the coming years and decades. And, even if development is not as rapid as imagined (Iansiti & Lakhani, 2017), the transformational power of foundational technologies is immense.

However, global institutions and national governments must work in concert with each other to mitigate fragmentation risks and ensure the inputs to algorithms used to develop AI models are complementary to human flourishing. The road to global coordination for blockchain standards and regulatory treatment is far more advanced and the same approach should be applied to AI.

As we look to the future, the synergy between AI and blockchain holds the key to unlocking sustainable technological advancements. Emerging trends such as decentralized finance, smart healthcare, and automated governance systems show promising applications. We encourage stakeholders in the AI and blockchain ecosystems to come together, leveraging their collective expertise to tackle the challenges and harness the opportunities discussed throughout this report.

As we look to the future, the synergy between AI and blockchain holds the key to unlocking sustainable technological advancements. It encourages us to reimagine the possibilities of digital innovation, grounded in the principles of fairness, transparency, and inclusivity. This journey towards integrating AI and blockchain not only propels us towards technological excellence but also ensures that our advancements contribute positively to society and the environment.

7. Bibliography

Agosto, A., Cerchiello, P., & Giudici, P. (2023). Bayesian learning models to measure the relative impact of ESG factors on credit ratings. International Journal of Data Science and Analytics. https://doi.org/10.1007/s41060-023-00405-9

Ahelegbey, D., & Giudici, P. (2022). NetVIX - a network volatility index for financial markets. Physica A: Statistical Mechanics and Its Applications, 594, 127017.

Aldasoro, I., Gambacorta, L., Giudici, P., & Leach, T. (2022). The drivers of cyber risk. Journal of Financial Stability, 60, 100989. https://doi.org/10.1016/j.jfs.2021.100989

Alameer, A., & Saleh, H. (2022). Reinforcement Learning in Quantitative Trading: A Survey. IEEE Access. https://doi.org/10.36227/techrxiv.19303853.v1

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives. https://doi.org/10.1257/jep.31.2.211

Alloghani, M. A. (2024). Artificial Intelligence and Sustainability. In Signals and communication technology. https://doi.org/10.1007/978-3-031-45214-7

Amirzadeh, R., Nazari, A., & Thiruvady, D. (2022). Applying Artificial Intelligence in Cryptocurrency Markets: A survey. Algorithms, 15(11), 428. https://doi.org/10.3390/a15110428

Arrieta, A. B., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Atsalakis, G. S. (2016). Using computational intelligence to forecast carbon prices. Applied Soft Computing, 43, 107–116. https://doi.org/10.1016/j.asoc.2016.02.029

Aziz, S., & Dowling, M. (2019). Machine learning and AI for risk management. Disrupting Finance: FinTech and Strategy in the 21st Century, 33–50.

Babaei, G. & Giudici, P. & Raffinetti, E. (2024). Safeaipackage: a Python package for a safe artificial intelligence. SSRN paper.

Bader, M. (2018). Disinformation in elections. Security And Human Rights, 29(1–4), 24–35. https://doi.org/10.1163/18750230-02901006

Baum, C., Chiang, J. H. Y., David, B., & Frederiksen, T. K. (2023). Eagle: Efficient privacy preserving smart contracts. In Lecture Notes in Computer Science (pp. 270–288). https://doi.org/10.1007/978-3-031-47754-6_16

Basly, S. (2024). Artificial intelligence and the future of decentralized finance. In Springer eBooks (pp. 175–183). https://doi.org/10.1007/978-3-031-49515-1_10

Ben-Sasson, E., Bentov, I., Horesh, Y., & Riabzev, M. (2018). Scalable, transparent, and post-quantum secure computational integrity. Cryptology ePrint Archive.

Borenstein, J., & Howard, A. M. (2020). Emerging challenges in AI and the need for AI ethics education. AI and Ethics, 1(1), 61–65. https://doi.org/10.1007/s43681-020-00002-7

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research, 81, 1–15. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Buterin, V. (2024). The promise and challenges of crypto + AI applications. Retrieved from https://vitalik.eth.limo/general/2024/01/30/cryptoai.html

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

Cerchiello, P., Giudici, P., & Nicola, G. (2017). Twitter data models for bank risk contagion. Neurocomputing, 264, 50–56. https://doi.org/10.1016/j.neucom.2016.10.101

Cihon, P., Maas, M. M., & Kemp, L. (2020). Should Artificial Intelligence Governance be Centralised?: Design Lessons from History. Social Science Research Network. https://doi.org/10.2139/ssrn.3761636

Dağlarli, E. (2020). Explainable artificial intelligence (xAI) approaches and deep meta-learning models. Advances and Applications in Deep Learning, 79. https://www.intechopen.com/chapters/72398

Demazeau, Y., & Müller, J. (1991). Decentralized A.I., 2. Elsevier.

Demia. (2024). Demia. Retrieved May 10, 2024 from https://www.demia.net/

Dinh, T. N., & Thai, M. T. (2018). AI and Blockchain: A disruptive integration. IEEE Computer, 51(9), 48–53. https://doi.org/10.1109/MC.2018.3620971

Eliaçik, A. B., & Erdoğan, N. (2015). User-weighted sentiment analysis for the financial community on Twitter. International Conference on Innovations in Information Technology. https://doi.org/10.1109/INNOVATIONS.2015.7381513

Ethembeni, D. (2021). Regulatory challenges of artificial intelligence systems. Springer Nature.

Espenan, N. P. (2023). Improving Voluntary Carbon Markets Through Standardization and Blockchain Technology. Wyoming Law Review, 23(1), 141–177. https://doi.org/10.59643/1942-9916.1473

European Commision (2024). AI Act in Shaping Europe's Digital Future. Retrieved May 10, 2024 from https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Fernandes, S., Sheeja, M. S., & Parivara, S. (2024). Potential of AI for a Sustainable, Inclusive, and Ethically Responsible Future. Fostering Multidisciplinary Research for Sustainability, 196.

Gangwal, A., Gangavalli, H. R. & Thirupathi, A. (2023). A survey of Layer-two blockchain protocols. Journal Of Network And Computer Applications, 209, 103539. https://doi.org/10.1016/j.jnca.2022.103539

Gelhaar, J., Gürpinar, T., Henke, M., & Otto, B. (2021). Towards a Taxonomy of Incentive Mechanisms for Data Sharing in Data Ecosystems. Pacific Asia Conference on Information Systems (PACIS) 2021, Volume: 25.

Giudici, P., & Abu-Hashish, I. (2019). What determines bitcoin exchange prices? A network VAR approach. Finance Research Letters, 28, 309–318.

Giudici, P., & Leach, T. (2022). Libra or Librae? Basket-based stable coins to mitigate foreign exchange volatility spillovers. Finance Research Letters, 44, 102504. https://doi.org/10.1016/j.frl.2023.104088

Giudici, P., & Polinesi, G. (2021). Network models to improve robot advisory portfolios. Annals of Operations Research, 313(2), 965–989. https://doi.org/10.1007/s10479-021-04312-9

Giudici, P., & Polinesi, G., & Spelta, A. (2022). Network models to improve robot advisory portfolios. Annals of Operations Research, 313(2), 965–989. https://doi.org/10.1007/s10479-021-04312-9

Giudici, P., & Raffinetti, E. (2022). Explainable AI methods in cyber risk management. Quality and Reliability Engineering International, 38(3). https://doi.org/10.1002/qre.2745

Giudici, P., & Raffinetti, E. (2023). SAFE Artificial Intelligence in Finance. Finance Research Letters, 56, 104088. https://doi.org/10.1016/j.frl.2023.104088

Giudici, P., & Venkatesh, V. (2020). Blockchain, adoption, and financial inclusion in India: Research opportunities. International Journal of Information Management, 52, 101936. https://doi.org/10.1016/j.ijinfomgt.2019.04.009

Giudici, P., Raffinetti, E., & Pagnottoni, P. (2022). Libra or Librae? Basket-based stable coins. Finance Research Letters, 44, 102504. https://doi.org/10.1016/j.frl.2023.104088

Giudici, P., Abu-Hashish, I., & Abu-Hashish, I. (2019). What determines bitcoin exchange prices? A network VAR approach. Finance Research Letters, 28, 309–318.

Giudici, P., Polinesi, G., & Spelta, A. (2021). Network models to improve robot advisory portfolios. Annals of Operations Research, 313(2), 965–989. https://doi.org/10.1007/s10479-021-04312-9

Goel, A. K., Bakshi, R. & Agrawal, K. K. (2022). Web 3.0 and Decentralized Applications. Materials Proceedings. https://doi.org/10.3390/materproc2022010008

Goodfellow, I., Bengio, Y., & Courville, A. (2020). Deep Learning. MIT Press.

Gopalakrishnan, P., & Ramaguru, R. (2019). Blockchain based waste management. International Journal of Engineering and Advanced Technology, 8(5), 2632–2635.

Große, N., Leisen, D., Gürpinar, T., Forsthövel, R., Schulze, R., Henke, M., & ten Hompel, M. (2020). Evaluation of (De-)Centralized IT technologies in the fields of Cyber-Physical Production Systems. Proceedings of the Conference on Production Systems and Logistics: CPSL, 377–387. https://doi.org/10.15488/9680

Gürpinar, T., Küpeli, O. (2023). Towards a definition of the industrial metaverse applied in context of the blockchain and web3 ecosystem. Blockchain and Cryptocurrency Conference.

Hamdan, A., Hassanien, A. E., Razzaque, A., & Alareeni, B. (2021). The Fourth Industrial Revolution: Implementation of artificial intelligence for growing business success. Springer Nature.

He, G., & Litterman, R. B. (2002). The intuition behind Black-Litterman model portfolios. Social Science Research Network. https://doi.org/10.2139/ssrn.334304

Hussain, A. A., & Al-Turjman, F. (2021). Artificial Intelligence and Blockchain: A review. Transactions on Emerging Telecommunications Technologies, 32(9). https://doi.org/10.1002/ett.4268

Iansiti, M., & Lakhani, K. R. (2017). The truth about blockchain. *Harvard Business Review.* Retrieved May 10, 2024, from https://hbr.org/2017/01/the-truth-about-blockchain

Internet Governance Forum; Programme Development; February 2024. https://intgovforum.org/en/filedepot_download/309/27171

Karunathilake, E. M. B. M., Le, A. T., Heo, S., Chung, Y., & Mansoor, S. (2023). The Path to Smart Farming: Innovations and Opportunities in Precision Agriculture. Agriculture, 13(8), 1593. https://doi.org/10.3390/agriculture13081593

Keary, T. (2023, December). Decentralized Artificial Intelligence (DAI). Techopedia. Retrieved from https://www.techopedia.com/definition/decentralized-ai-dai

Knight, W. (2021). The Dark Secret at the Heart of AI. MIT Technology Review. https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/

Knott, B., Venkataraman, S., Hannun, A., Sengupta, S., Ibrahim, M., & Van Der Maaten, L. (2021). CryPTeN: Secure Multi-Party Computation Meets Machine Learning. arXiv (Cornell University), 34. https://arxiv.org/abs/2109.00984

Kudelski Group. (n.d.). QTAKE Delivers Industry-First by Integrating Forensic Watermarking at Camera. Retrieved April 12, 2024, from https://www.nagra.com.

Küpeli, O., & Gürpinar, T. (2023). Towards a definition of the industrial metaverse applied in context of the blockchain and web3 ecosystem. Blockchain and Cryptocurrency Conference.

Liao, W., Lu, X., Fei, Y., Gu, Y. & Huang, Y. (2024). Generative AI design for building structures. Automation in Construction, 157, 105187. https://doi.org/10.1016/j.autcon.2023.105187

Lifong, Z., Yan, Z., & Raimo, K. (2016). A review of Homomorphic Encryption and its applications. In Mobimedia 2016: Proceedings of the 9th EAI International Conference on Mobile Multimedia Communications (pp. 97–106).

Luo, B., Zhen, Z., Wang, Q., Ke, A. H. P., Lu, S., & He, B. (2023). AI-powered fraud detection in Decentralized Finance: A project Life cycle perspective. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2308.15992

Ma, X., Wang, C., & Chen, X. (2021). Trusted data sharing with flexible access control based on blockchain. Computer Standards & Interfaces, 78, 103543. https://doi.org/10.1016/j.csi.2021.103543

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017 (p. 54). PMLR.

Meeco Group Pty Ltd. (2023, Juli). EU Digital Identity: EBSI, EUDI and Verifiable Credentials for Education. www.meeco.me. Abgerufen am 13. Juni 2024, von https://www.meeco.me/blog/ebsi-verifiable-credentials-for-education

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

Nah, F. F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research, 25(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). Bitcoin and Cryptocurrency Technologies. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691171692/bitcoin-and-cryptocurrency-technologies

Nassar, M., Salah, K., Rehman, M. H. U., & Svetinovic, D. (2019). Blockchain for explainable and trustworthy artificial intelligence. WIREs Data Mining and Knowledge Discovery, 10(1). https://doi.org/10.1002/widm.1340

Naqash, M. T., Syed, T. A., Alqahtani, S. S., Siddiqui, M. S., Alzahrani, A., & Nauman, M. (2023). A Blockchain Based Framework for Efficient Water Management and Leakage Detection in Urban Areas. Urban Science, 7(4), 99. https://doi.org/10.3390/urbansci7040099

Nartey, J. (2024). Decentralized Finance (DeFi) and AI: Innovations at the Intersection of Blockchain and Artificial Intelligence. Social Science Research Network. https://doi.org/10.2139/ssrn.4781328

Nassar, M., Salah, K., Rehman, M. H. U., & Svetinovic, D. (2019). Blockchain for explainable and trustworthy artificial intelligence. WIREs Data Mining and Knowledge Discovery, 10(1). https://doi.org/10.1002/widm.1340

OECD (2023). Updated definition of Artificial Intelligence. https://oecd.ai/en/wonk/ai-system-definition-update

Paola Cerchiello, Paolo Giudici, Giancarlo Nicola (2017). Twitter data models for bank risk contagion. Neurocomputing, 264, 50-56. https://doi.org/10.1016/j.neucom.2016.10.101

Paramanik, R. N., & Singhal, V. (2020). Sentiment analysis of Indian stock market volatility. Procedia Computer Science, 176, 330–338. https://doi.org/10.1016/j.procs.2020.08.035

Rabetti, D. (2023). Auditing decentralized Finance (DEFI) protocols. Social Science Research Network. https://doi.org/10.2139/ssrn.4458298

Raj, R., Rimal, B. P., & Maier, M. (2020). Machine Learning for Blockchain and Blockchain for Machine Learning. IEEE Communications Surveys & Tutorials, 22(3), 2215–2255.

Rawls, J. (1958). Justice as Fairness. ˜The œPhilosophical Review, 67(2), 164. https://doi.org/10.2307/2182612

Rocha, D. P., Kinkelin, H. & Rezabek F. (2022). Secure Data Marketplaces. Chair of Network Architectures and Services, Department of Informatics, Technical University of Munich, Germany. In Seminar IITM SS 22, Network Architectures And Services, November 2022 (p. 47). https://doi.org/10.2313/NET-2022-11-1_09

Russell, S., & Norvig, P. (2010). Artificial Intelligence: A Modern Approach. Prentice Hall. https://aima.cs.berkeley.edu/

Sadman, N., Rahman, A., & Gupta, K. D. (2021). Promise of AI in Defi, A Literary Analysis. Digital. https://doi.org/10.20944/preprints202110.0136.v2

Salah, K., Rehman, M. H. U., Nizamuddin, N., & Al-Fuqaha, A. (2019). Blockchain for AI: Review and open research challenges. IEEE Access, 7, 10127–10149. https://doi.org/10.1109/access.2018.2890507

Santana, C. & Albareda, L. (2022). Blockchain and the emergence of Decentralized Autonomous Organizations (DAOs): An integrative model and research agenda. Technological Forecasting & Social Change/Technological Forecasting And Social Change, 182, 121806. https://doi.org/10.1016/j.techfore.2022.121806

Sartor, G., & Lagioia, F. (2020). The impact of the General Data Protection Regulation (GDPR) on artificial intelligence.

Schuetz, S., & Venkatesh, V. (2020). Blockchain, adoption, and financial inclusion in India: Research opportunities. International Journal Of Information Management, 52, 101936. https://doi.org/10.1016/j.ijinfomgt.2019.04.009

Schneier, B. (2008). The psychology of security. Computer security, 26(5), 20–25. https://doi.org/10.1109/ms.2008.137

Schneier, B. (2015). Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World. W. W. Norton & Company.

Skinner, A., & Singh, N. (2021). Edge-ai: Decentralizing Artificial Intelligence. ACM.

Sivarethinamohan, R., & Sujatha, S. (2021). Unraveling the potential of artificial intelligence-driven blockchain technology in environment management. In Advances in Mechanical Engineering: Select Proceedings of CAMSE 2020 (pp. 693–700). Springer Singapore.

Solove, D. J. (2024). Artificial Intelligence and Privacy. Social Science Research Network. https://doi.org/10.2139/ssrn.4713111

Sullivan, Y., & Wamba, S. F. (2024). Artificial intelligence and adaptive response to market changes: A strategy to enhance firm performance and innovation. Journal of Business Research, 174, 114500.

Tapscott, D., & Tapscott, A. (2016). Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World. https://itig-iraq.iq/wp-content/uploads/2019/05/Blockchain_Revolution.pdf

Taherdoost, H. (2024). Blockchain Integration and Its Impact on Renewable Energy. Computers, 13(4), 107. https://doi.org/10.3390/computers13040107

Tyagi, K. (2023a). Deepfakes, Copyright and Personality Rights an Inter-Disciplinary Perspective. In Law and Economics of the Digital Transformation (S. 191–210). https://doi.org/10.1007/978-3-031-25059-0_9

Tyagi, K. (2023b). A global blockchain-based agro-food value chain to facilitate trade and sustainable blocks of healthy lives and food for all. Humanities & Social Sciences Communications, 10(1). https://doi.org/10.1057/s41599-023-01658-2

Vaghani, A., Gürpinar, T., & Grosse, N. (2022). A Taxonomy Characterizing Blockchain-Empowered Services for the Metaverse. Blockchain and Cryptocurrency Conference.

Verbeek, A., & Lundqvist, M. (2021). Artificial Intelligence, Blockchain and the Future of Europe: How Disruptive Technologies Create Opportunities for a Green and Digital Economy: Main Report. European Investment Bank. https://www.eib.org/attachments/thematic/artificial_intelligence_blockchain_and_the_future_of_europe_report_en.pdf

Vitalik. (2024). The promise and challenges of crypto + AI applications. Retrieved January 30, 2024, from https://vitalik.eth.limo/general/2024/01/30/cryptoai.html

Wright, A., & De Filippi, P. (2018). Blockchain and the law: the rule of code. Harvard University Press. https://larc.cardozo.yu.edu/faculty-books/108/

Xu, J., Liu, S., Zhao, H., & Yang, M. (2022). A brief survey of federated learning. Complexity, 2022, 1–14. https://doi.org/10.1155/2022/75195l1

Yang, S., Sun, L., Yan, Q., & Wan, J. (2020). Review of Big Data and Artificial Intelligence Enabled Smart Healthcare Applications. IEEE Access, 8, 22314–22334.

Yassine, A., Usman, M., Javed, A., Faisal, M., Afzal, H., & Saleem, K. (2020). Decentralized healthcare systems with blockchain smart contracts: A case study of COVID-19. Information Technology & People.

Yi, X., Paulet, R. & Bertino, E. (2014). Homomorphic encryption. In SpringerBriefs in computer science (S. 27–46). https://doi.org/10.1007/978-3-319-12229-8_2

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man, and Cybernetics, (1), 28–44. https://doi.org/10.1109/tsmc.1973.5408575

Zadeh, L. A. (1996). Fuzzy logic = computing with words. Fuzzy Systems, IEEE Transactions, 4(2), 103–111. https://doi.org/10.1109/91.493904