Fundamentals of Decentralized AI

FEBRUARY 2025
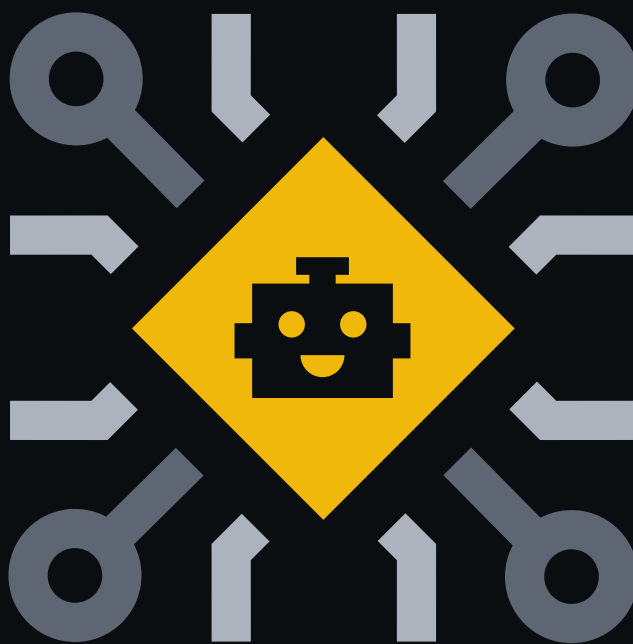
Joshua Wong

Table of Contents

01 / Key Takeaways

- Software has eaten the world; now artificial intelligence (AI) is eating software. Most of the AI of today however, appears to be developed and controlled by centralized Big Tech companies.

- Since Bitcoin's launch in 2008, crypto technology has been paving the way for decentralized networks that could help to democratize AI.

- Decentralized AI (DeAI) goes beyond the AI agents currently swarming crypto Twitter, many of which rely on centralized Large Language Models (LLMs). The DeAI movement aims to build networks and infrastructure for creating AI models in a decentralized way.

- AI fundamentally consists of three components: data, compute, and algorithms. Even before the public launch of ChatGPT in 2022, various Decentralized Physical Infrastructure (DePIN) projects had already been building the necessary infrastructure required to offer these components in a decentralized manner, forming the building blocks of Decentralized AI (DeAI).

- As the AI race amongst the global superpowers begins to accelerate beyond individual control, distributed networks may begin to attract attention as a means of consolidating resources and AI-related power in a decentralized manner.

02 / Introduction

Software has eaten the world; now AI is eating software. As AI that is able to iterate on its own software begins to emerge, this revolutionary technology seems poised to become the most rapid and impactful phenomenon to ever propagate through human civilization.

"Software is eating the world. And AI is eating software."

—Jensen Huang, CEO of NVIDIA (2017)

AI has quickly captured the attention of individuals, companies, and governments, driving rapid innovation and raising concerns about the centralization of AI power. Major organizations appear to be scrambling to secure compute power and data for training their models. Meanwhile blockchain technology, initially designed to decentralize financial power, offers a timely solution to prevent the centralization of AI systems. **By providing transparency, immutability, defined ownership, and an adversarial testing environment, blockchains can enhance AI systems** just as they have revolutionized financial systems.[1]

"There is a significant chance… that AI applications will be crypto's raison d'être."

—Matthew Sigel, Head of Digital Assets Research at VanEck

Long seen as a transformative technology in search of its most impactful applications, blockchains may have finally found their raison d'être: combating the centralization of AI systems. **Similar to money, AI is a technology poised to infiltrate nearly every aspect of human life.** However, centralized AI could lead to concerns over control, transparency, and bias, as a small group of entities may have disproportionate influence over AI behavior and decision-making.

This report explores the concerns related to an overly centralized global AI system and how blockchain technology could mitigate these risks as AI becomes increasingly integrated into our daily lives.

2.1 An Overview of the AI industry

The modern AI industry appears to be led by a small number of models, developed by companies that have made or raised billions of dollars to spend on acquiring compute power and accessing the largest data sets. In the short time since ChatGPT launched to the public in 2022, **we've already seen significant developments in the capabilities of these LLMs**, with many of them moving from 'text only' capabilities to becoming multimodal.

Figure 1: LLMs have already made significant advancement since the launch of ChatGPT in 2022

Owned by	2022-2023	2025
Deepseek	NA	Janus Pro • Multimodal (text, images) • Advanced multimodal Reasoning (capable of multi-step problem-solving and nuanced analysis, using both text and images)
Anthropic	Claude • Not multimodal (text only) • Limited contextual understanding (difficulty with complex conversation) • No tool usage	Claude 3.5 • Multimodal (text, audio, images) • Enhanced contextual understanding (coherent during long interactions) • Experimental computer usage capability
Google	Google Bard • Not multimodal (text only) • Fair reasoning • Limited contextual understanding (difficulty with complex conversation) • Limited real time data integration • Low personalization (limited adaptability)	Gemini 2.0 Flash • Multimodal (text, audio, images) • Advanced Reasoning (capable of multi-step problem-solving and nuanced analysis) • Enhanced contextual understanding (coherent during long interactions) • Real-time data integration (from Google Search) • Advanced Personalization (user context)
Meta	Llama 1 • Not multimodal (text only) • Fair reasoning • Limited contextual understanding (difficulty with complex conversation) • No API access	Llama 3.3 • Text-based (Llama 3.2 was multimodal) • Advanced Reasoning (capable of multi-step problem-solving and nuanced analysis) • Enhanced contextual

			• understanding (coherent during long interactions) • API access (tools for model and agent development)
Microsoft	Phi-1 • Not multimodal (text only) • Fair reasoning (limited to coding tasks) • Focused training (smaller, coding-focused data set)		Phi-4 • Multimodal (text, audio, images) • Advanced Reasoning (capable of multi-step problem-solving and nuanced analysis) • Comprehensive training (diverse data)
OpenAI	GPT 3.5 • Not multimodal (text only) • Fair reasoning (scored high on SAT but bottom 10% on bar examination) • Standard API access (for text generation)		OpenAI o1 • Multimodal (text and images) • Advanced Reasoning (top 10% on bar examination) • Enhanced contextual understanding (coherent during long interactions) • Advanced API access (supports multimodal inputs)

Source: Mckinsey, Binance Research

As capital and resources pour into AI development, the advancement of models owned by large organizations may easily outpace smaller competitors. **Given AI's potential for rapid self-improvement, the risk of centralizing AI power in the hands of a few may be a real concern.**

Recently gaining mainstream attention, blockchain might just offer a solution to AI centralization. While major companies have been investing billions in AI research, the blockchain industry has been **busy building decentralized networks and infrastructure that could enable a free and open digital future**.

2.2 The Enclosure of the Internet Commons

A significant portion of internet activity data is collected and managed by major tech companies, raising concerns about centralized control and data privacy. Similarly, **access to critical digital resources - such as compute power and data storage - is often concentrated** among leading cloud service providers with the financial capacity to scale their infrastructure globally. Much like historical examples where public resources were consolidated for profit, concerns have emerged about how access to data and compute power is increasingly concentrated among a few major players.[2]

As we accelerate into the age of AI, it is these very resources that may fuel the production of the most powerful AI models. Blockchain technology - by giving the world a means of coordinating people and machines in a decentralized, trustless manner - has the potential to counterbalance the increasing concentration of the Internet commons in the hands of Big Tech.
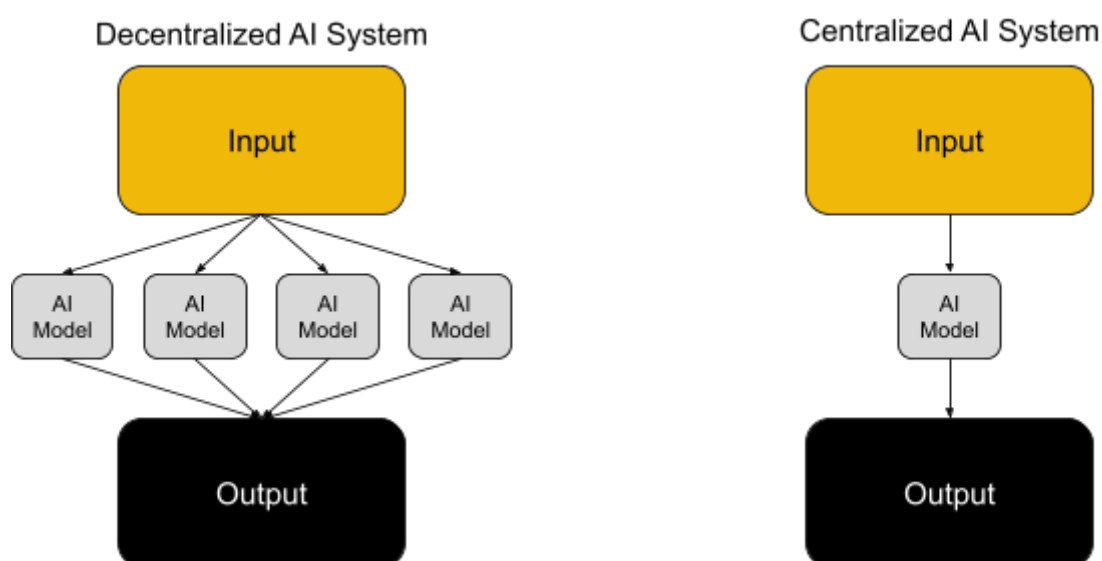
03 / What is Decentralized AI?

The topic of decentralized AI is certainly a complex one, especially given that we're only now beginning to grapple with how AI itself (much less its decentralized version) might look and operate in the world.

To illustrate what the decentralized AI landscape may look like in the future, we may imagine **a system that leverages multiple AIs, controlled and trained by multiple different parties, which work together to respond with valuable outputs to a user's input**. This would contrast a centralized AI, which leverages a single model trained and controlled by a centralized organization (or perhaps also multiple models trained and controlled by a centralized organization).

Such a system would seamlessly combine the capabilities of multiple (perhaps specialized) models into the user experience of utilizing a single model. Models in this system would also compete to produce the best outputs, which could significantly drive down costs to the end user.

Figure 2: Future DeAI systems could leverage multiple separately trained models to produce equal or better outputs

Source: Bittensor, Binance Research

An article published by VanEck provides a perspective on the role decentralized AI infrastructure could play in a future where AI is integral to the global economy, and compute and storage are viewed as public utilities like power generation. In such an AI-driven economy, **the majority of the "base load" will be supplied by GPU cloud Hyperscalers** such as Amazon and Google, which may capture up to 80% of the market. **Blockchain-based server infrastructure will address specialized needs and serve as "peaking" providers during periods of high demand**. For producers of bespoke AI models, crypto storage and computing offer advantages such as on-demand service, shorter SLA lock-ins, customized environments, and improved latency sensitivity.
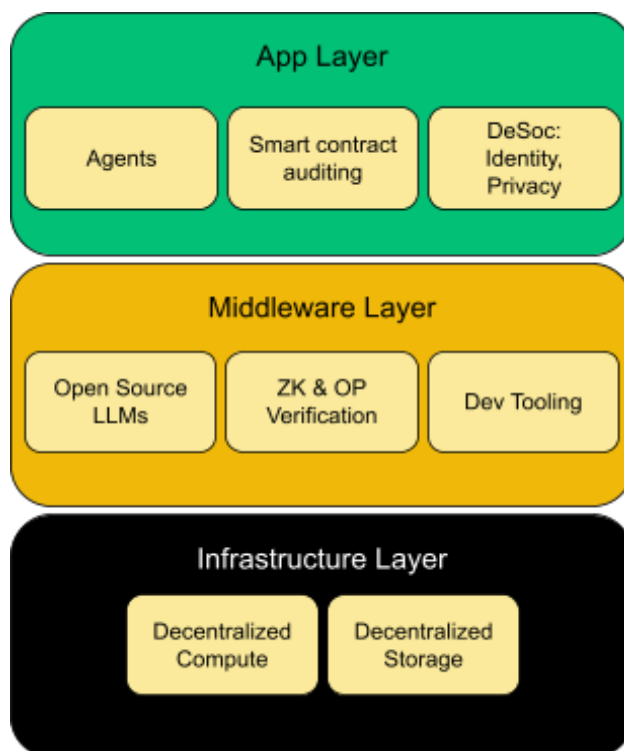
Decentralized GPUs can seamlessly integrate with decentralized AI models within smart contracts, enabling permissionless scaling of compute needs. Consequently, **blockchain-provisioned GPUs, comparable to Uber/Lyft for AI compute, could capture a portion of the non-Hyperscaler AI infrastructure market** (estimated by Vaneck to be ~20%).

3.1 Beneath the Agents

While AI Agents have only entered the crypto scene in full force in the last year or so, there are a few AI-adjacent infrastructure projects that have been building in the space for a number of years now. As the global fervor surrounding AI heated up, some of these **AI and AI-adjacent decentralized infrastructure projects have grown to become some of the largest** in the crypto space.

The AI agents gaining significant traction on crypto-related social media platforms today represent the top layer of the DeAI stack. They are like the end user applications that leverage the underlying technology stack. **Most, if not all, of the AI agents in the crypto space today utilize LLMs developed by centralized organizations**.

Figure 3: A simplified look at the decentralized AI stack

App Layer
- Agents
- Smart contract auditing
- DeSoc: Identity, Privacy

Middleware Layer
- Open Source LLMs
- ZK & OP Verification
- Dev Tooling

Infrastructure Layer
- Decentralized Compute
- Decentralized Storage

Source: Adapted from Delphi Digital, Binance Research

DeAI extends beyond the use of centrally-developed models operating on decentralized blockchains. **AI agents are part of the App Layer, which just makes up the tip of the DeAI iceberg**. Below agents, we can find frameworks which help developers build and

deploy these agents. It's only one level below that where we get into the infrastructure guts of the decentralized AI stack. At this level, blockchain-based coordination layers come into play, incentivizing decentralized parties to provide compute resources for tasks such as AI model training and data storage. These foundational elements form the basic infrastructure that enables the development of decentralized AI.

This report will focus on the infrastructure layer, which is responsible for making available the basic components required to create AI models in a decentralized manner.

04 / DeAI Basic Components

The complexity of AI can be distilled into three fundamental components - data, compute, and algorithms. It is, in essence, these three inputs that give us the output that is AI.

"AI has three fundamental components—data, compute, and algorithms"

—Wendy Wong, Professor of Political Science at UBC

We'll frame this section of the report around these three fundamental components, and explore how various projects are building the necessary infrastructure to facilitate and incentivize their creation and supply in a decentralized manner.

Figure 4: DeAI-related projects have grown to become some of the largest in the crypto space by token market capitalization, covering different fundamental components

Project	Data	Compute	Algorithms	Project Focus Area
NEAR Protocol	✔	✔	✘	General purpose blockchain capable of handling data and compute functions
Bittensor	✔	✔	✔	Specifically designed for DeAI. Focus on incentivizing development and deployment of AI models
Internet Computer Protocol	✔	✔	✔	General purpose decentralized cloud computing, including data storage and computation
Filecoin	✔	✘	✘	Decentralized data storage solution
Render	✘	✔	✘	Decentralized GPU compute, primarily for graphics rendering
Artificial Superintelligence Alliance (formerly Fetch.ai)	✔	✔	✔	Collaborative effort between Fetch.ai, SingularityNET, Ocean Protocol to further DeAI
The Graph	✔	✘	✘	Decentralized data storage solution

Arweave	✔	✘	✘	Decentralized storage solution
AIOZ	✔	✔	✔	Integrates AI into its decentralized content delivery network. Includes AI-powered services and an AI model marketplace
Akash Network	✘	✔	✘	Decentralized cloud computing resources. Not specifically designed for AI/ML

Source: Coinmarketcap, Binance Research as of February 2025 (non-comprehensive list, projects were selected from the top few by market capitalization under Coinmarketcap's AI & Big Data category)

4.1 Data

While AI agents have recently gained attention in the cryptocurrency sector, AI-adjacent decentralized infrastructure projects have been expanding for years. Some of the largest by token market capitalization focus on decentralized data storage. **Data is essential for building AI, and before decentralized storage, it relied on centralized providers** like AWS, Google Cloud, and Microsoft Azure, which have **risks like single points of failure, data breaches, and censorship.**

Decentralized storage solutions like Filecoin and Arweave address these issues by **incentivizing decentralized data storage, ensuring data integrity, availability, and redundancy** through blockchain and cryptographic security. These networks establish a decentralized framework for data storage and retrieval, potentially laying the groundwork for truly decentralized AI models. This infrastructure enhances data security and accessibility, paving the way for AI free from centralized control.

Moreover, **data stored on these networks is, by default, available to the public**. This infrastructure **converts data into a public good**, accessible to anyone with an Internet connection. This openness fosters the growth of decentralized AI, providing a solid and transparent foundation for AI development and innovation.

Filecoin: Incentivizing Decentralized Storage

Filecoin is an example of a project that is advancing the field of decentralized data storage, contributing to the development of decentralized AI. By building on top of the **Interplanetary File System (IPFS)**, Filecoin **introduces an incentivized storage layer that enhances data management and sharing**. IPFS uses content addressing, identifying files by their cryptographic hash, to facilitate efficient and censorship-resistant data sharing. Protocol Labs, the developers of both IPFS and Filecoin, recognized the need for a sustainable economic model and created a marketplace where storage providers are compensated with FIL tokens for reliably storing data.

Filecoin employs consensus mechanisms called Proof-of-Spacetime (PoSt) and Proof-of-Replication (PoRep) to verify storage claims and ensure data integrity.

- **Proof of Spacetime (PoSt)**: a consensus mechanism used by Filecoin to verify that storage providers are continuously storing the promised data over a specified period. It ensures that the data remains available and intact by requiring storage providers to submit regular proofs that they are still holding the data.

- **Proof-of-Replication (PoRep)**: a mechanism that ensures a storage provider has created unique copies of the data they claim to store. It verifies that each copy is physically distinct and not just a duplicate, thereby preventing storage providers from cheating the system by claiming multiple rewards for a single copy of data.

Tokenomic incentives improve data availability and redundancy within the IPFS network, promoting a more reliable and sustainable storage ecosystem. By incentivizing storage providers, Filecoin offers a balanced solution for decentralized and resilient data management, supporting the growth of decentralized AI.

Arweave: Permanent Data Storage

Similar to Filecoin, Arweave is a decentralized data storage network. However, instead of Filecoin's recurring payment for continued data storage, **Arweave utilizes a one-time upfront payment for permanent data storage**. This payment funds the long-term rewards for Arweave miners, incentivizing the indefinite storage of data on Arweave's decentralized network of computers.
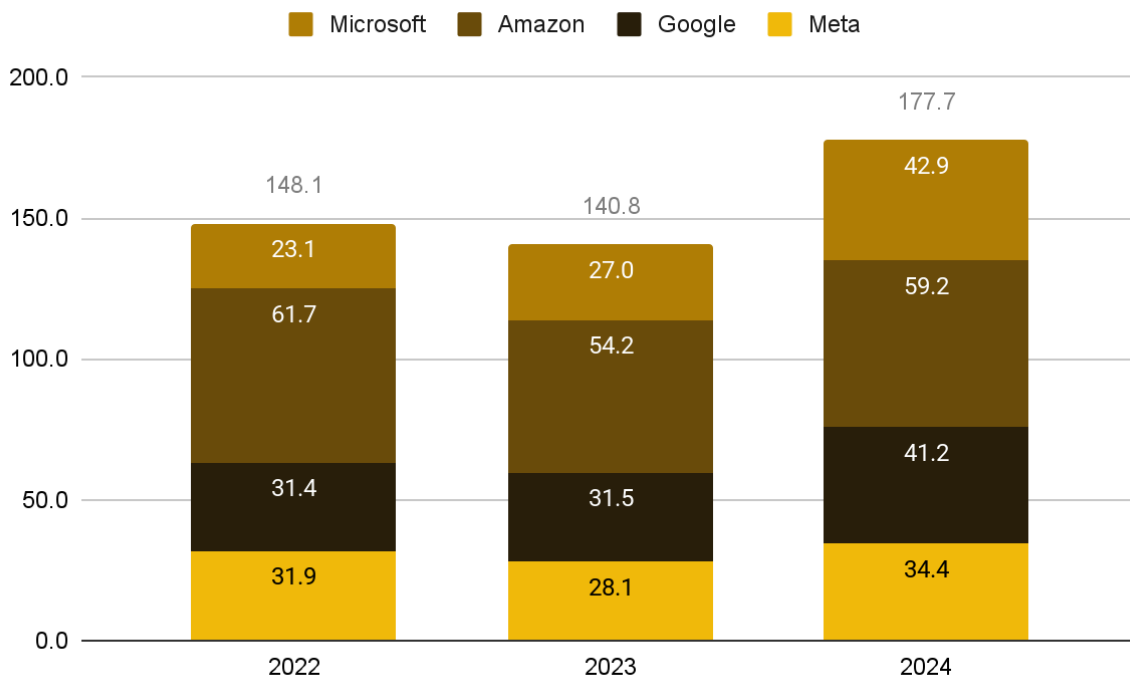
The differentiating feature of Arweave is its commitment to data permanence. Once data is stored on the network, it cannot be altered or deleted. **This permanence could prove particularly useful for decentralized AI research, as it guarantees the integrity and consistency** of the data used for training and decision-making processes, providing a reliable and replicable foundation for AI modeling.

By integrating permanent storage with a sustainable economic framework, Arweave delivers a compelling solution for decentralized data archiving. Through its enhancement of data integrity and accessibility, Arweave's permanent data storage may form the foundation of future key DeAI research and development.

4.2 Compute

The global race to acquire compute power appears to be heating up. Big Tech companies are leveraging their cash reserves, bolstered by their dominant positions in the digital economy, to acquire cloud infrastructure for both data and compute power.[3]

Figure 5: Big Tech has increased spending on compute hardware for AI training by 26% between 2023 and 2024

Source: Delphi Digital, Binance Research

Where data forms the foundation of AI, compute power is the essential resource that enables AI models to be trained on said data. Before cryptography allowed us to create and incentivize decentralized actors to commit compute power to a global network in order to achieve a specific task, we would solely rely on renting compute power from the organizations that have deep enough pockets to purchase vast amounts of computing power.

Today, projects like Render and AIOZ are among several initiatives **developing distributed networks to coordinate the decentralized contribution of compute power for the purpose of AI training.** Similar to the decentralized data storage networks, decentralized computing networks **transform computational power into a publicly accessible commodity**. This opens the possibility for a wider variety of parties to leverage said compute to train new AI models.

Render: Incentivizing Idle Compute

Initially designed to rent out dormant GPU power for rendering graphics, the Render Network of today also supports AI computing, **allowing GPU power to be rented for training generative AI models**. Using blockchain technology, Render Network distributes tasks across a global node network, moving away from centralized models. This enhances the efficiency and scalability of AI applications by providing a robust, distributed compute environment.

Render Network **aims to democratize access to high-performance computing resources**, making them more accessible and cost-effective for AI developers. Their system incentivizes the coordinated use of GPUs worldwide for tasks like machine learning. The **Proof of Render (PoR) consensus mechanism verifies and rewards computational tasks**, ensuring efficient performance. By **turning computation power into a publicly accessible commodity**, Render Network has the potential to optimize AI workflows, reduce latency, and support the growth and training of decentralized AI models.

AIOZ: Edge AI Compute

As a decentralized Content Delivery Network (CDN) utilizing blockchain technology, **AIOZ focuses on enhancing computational and storage resources for AI applications**. The project distributes storage and compute capabilities across a global network of nodes, moving away from traditional centralized CDN models. This **distributed architecture is designed to address latency issues**, which are important for AI training and real-time response scenarios. By positioning compute power closer to data sources, AIOZ aims to optimize AI workflows across the globe.

While Render is dedicated exclusively to decentralized compute, **AIOZ aims to establish a comprehensive Web3 AI infrastructure** that democratizes access to all three fundamental components of decentralized AI (DeAI): compute, storage, and algorithms. According to their **AIOZ W3AI Vision Paper** published in 2024, this ambitious goal positions AIOZ as a noteworthy project, being one of the few initiatives addressing the entire DeAI stack within a single decentralized network.

4.3 Algorithms

The final basic component of AI is algorithms, which **currently requires the expertise of human engineers** . OpenAI, for example, is estimated to spend a significant **$1.5 billion annually on labor costs.**

To allow for the creation of decentralized AI systems to exist, **new methods are being explored to incentivize human talent to design the algorithms needed** for training new AI models, without relying solely on centralized organizations with large budgets.

Bittensor: AI Model Distribution

Bittensor is another example of a decentralized network designed to facilitate and incentivize multiple aspects of development and deployment of AI models (data storage, compute, model creation etc.). Launched in 2019, Bittensor introduced its initial mainnet, "Kusanagi," in January 2021, and later transitioned to its own blockchain, "Nakamoto," in March 2023.

A key functionality of Bittensor is providing a decentralized **platform where machine learning models and computational resources can be exchanged securely and transparently**, thanks to blockchain technology. This setup encourages collaboration and innovation within the AI community.

The network's unique Proof of Intelligence (PoI) consensus mechanism plays a crucial role in this ecosystem. It **evaluates and rewards participants based on the quality and utility of their AI models, thereby promoting the development and sharing of high-quality models** in a collaborative environment.

Bittensor's TAO tokenomics, with its limited supply and halving events similar to Bitcoin, are designed to shape the network's economic model and incentivize participation. This approach aims to **enhance the distribution of computational resources and model development** within the AI sector.

By offering a decentralized platform for building and proliferating AI models, Bittensor provides **an alternative to the traditional reliance on large tech companies** with significant financial resources. This initiative is paving the way for a more distributed and democratized approach to AI development.

05 / Closing Thoughts

Where blockchains were invented and deployed (in the form of Bitcoin initially) as a response to the perceived over-centralization of an established financial system, **progress on DeAI projects seems to be happening simultaneously to humanity's deployment of AI technology**. If nothing else, this displays the acceleration of humanity's technological advancement and the potential to develop the technological means to democratically distribute a revolutionary technology, whilst that revolutionary technology is itself being built.

This report concludes by noting the **growing intersection between blockchain technology and artificial intelligence in the form of DeAI**. While AI has long been at the forefront of research and public discourse, blockchain technology has experienced a more gradual path to mainstream acceptance, initially emerging from niche online communities. As AI advancements progress rapidly, blockchain technology is emerging in parallel as a potential counterbalance to concerns around centralization. As the global AI race intensifies, **blockchains and their distributed networks may provide promising avenues for mitigating the risks of centralized AI control**.

06 / References

1. https://www.vaneck.com/us/en/blogs/digital-assets/matthew-sigel-vanecks-crypto-ai-revenue-predictions-by-2030/

2. https://www.linuxfoundation.org/research/decentralized-internet

3. https://members.delphidigital.io/feed/deai-compressed

07 / New Binance Research Reports

Monthly Market Insights - February 2025 <u>Link</u>
A summary of the most important market developments, interesting charts and upcoming events

From Challenges to Opportunities: How DeSci Reimagines Science <u>Link</u>
An introduction to DeSci

About Binance Research

Binance Research is the research arm of Binance, the world's leading cryptocurrency exchange. The team is committed to delivering objective, independent, and comprehensive analysis and aims to be the thought leader in the crypto space. Our analysts publish insightful thought pieces regularly on topics related but not limited to, the crypto ecosystem, blockchain technologies, and the latest market themes.
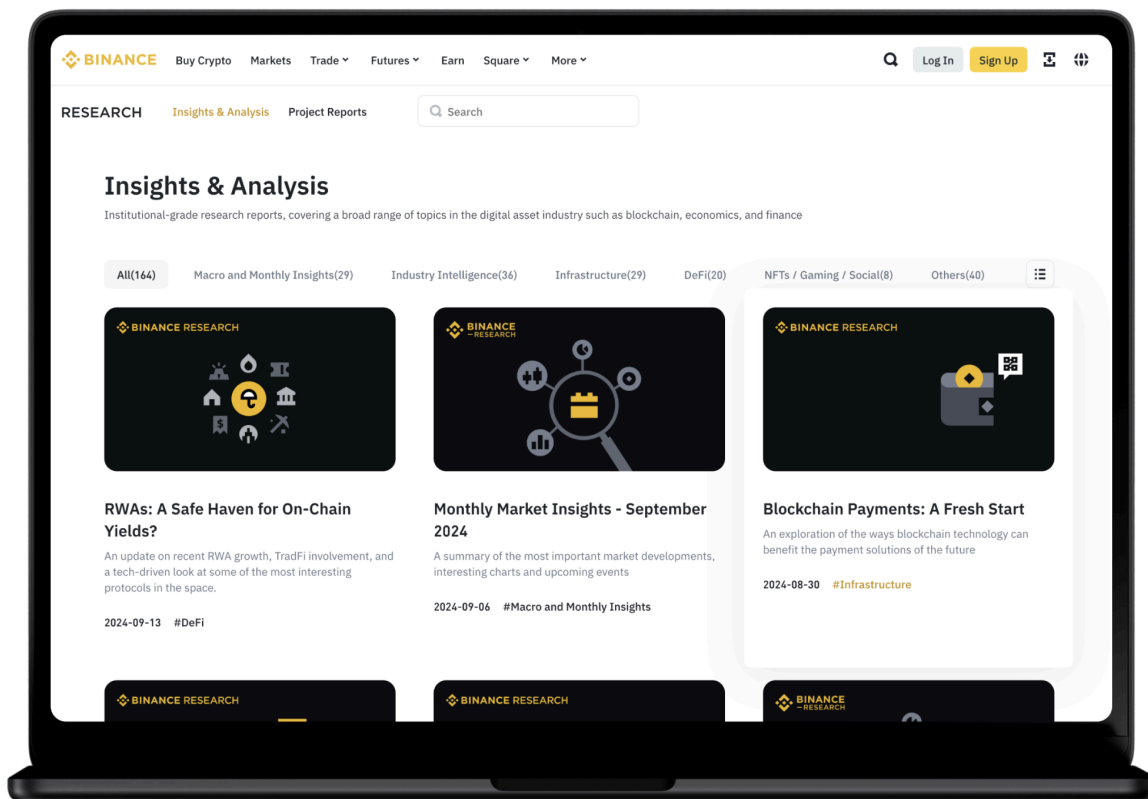
Joshua Wong
Macro Researcher

Joshua is currently working for Binance as a Macro Researcher. He has been involved in the cryptocurrency space since 2019. Prior to joining Binance, he worked as a product manager at a Web3 fintech startup, and a market analyst at a DeFi startup. He holds a Bachelor of Laws (LLB) from Durham University.

Resources

Binance Research **Link**

 Share your feedback **here**

GENERAL DISCLOSURE: This material is prepared by Binance Research and is not intended to be relied upon as a forecast or investment advice, and is not a recommendation, offer or solicitation to buy or sell any securities, cryptocurrencies or to adopt any investment strategy. The use of terminology and the views expressed are intended to promote understanding and the responsible development of the sector and should not be interpreted as definitive legal views or those of Binance. The opinions expressed are as of the date shown above and are the opinions of the writer, they may change as subsequent conditions vary. The information and opinions contained in this material are derived from proprietary and non-proprietary sources deemed by Binance Research to be reliable, are not necessarily all-inclusive and are not guaranteed as to accuracy. As such, no warranty of accuracy or reliability is given and no responsibility arising in any other way for errors and omissions (including responsibility to any person by reason of negligence) is accepted by Binance. This material may contain 'forward looking' information that is not purely historical in nature. Such information may include, among other things, projections and forecasts. There is no guarantee that any forecasts made will come to pass. Reliance upon information in this material is at the sole discretion of the reader. This material is intended for information purposes only and does not constitute investment advice or an offer or solicitation to purchase or sell in any securities, cryptocurrencies or any investment strategy nor shall any securities or cryptocurrency be offered or sold to any person in any jurisdiction in which an offer, solicitation, purchase or sale would be unlawful under the laws of such jurisdiction. Investment involves risks. For more information, please click here.