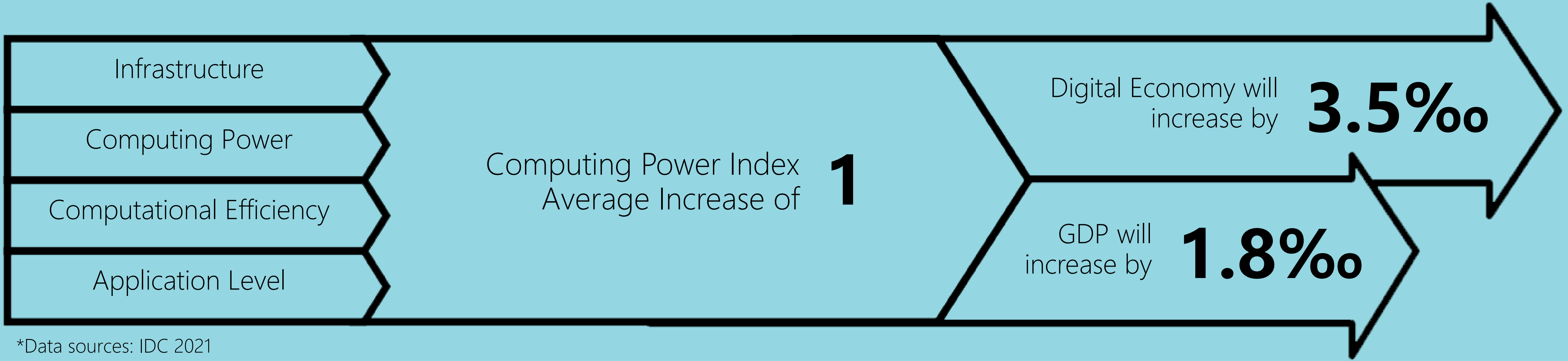
Matrix AI

Network As Computers

Decentralied Computing Network

MatrixAI

Computing | An area closely related to the world economy

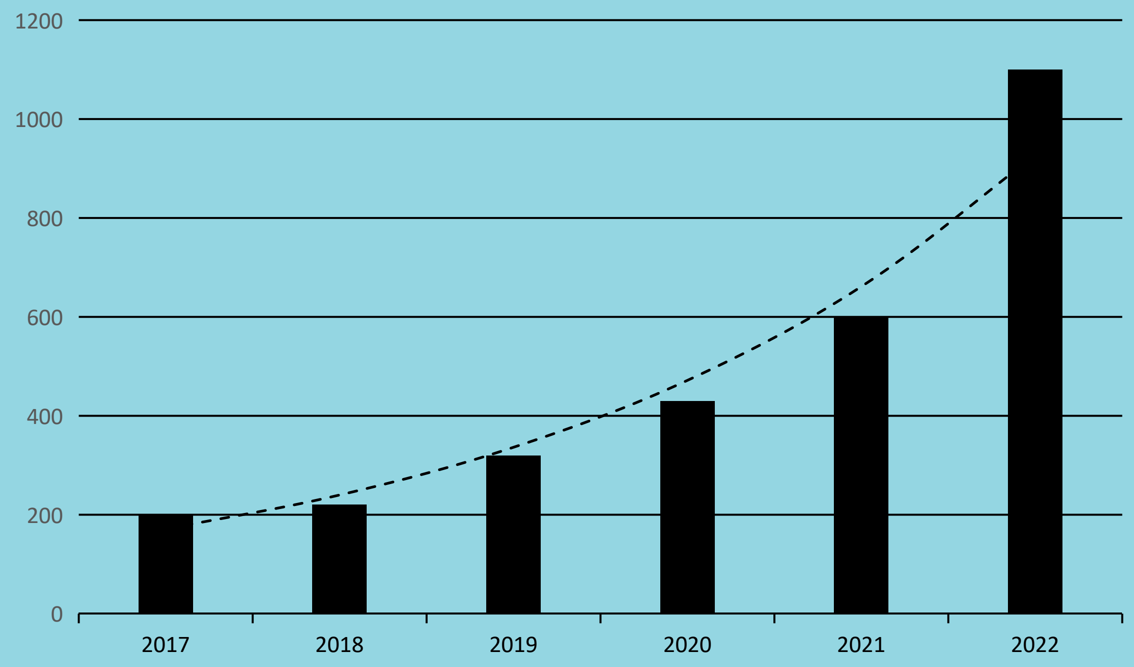


*Data sources: IDC 2021

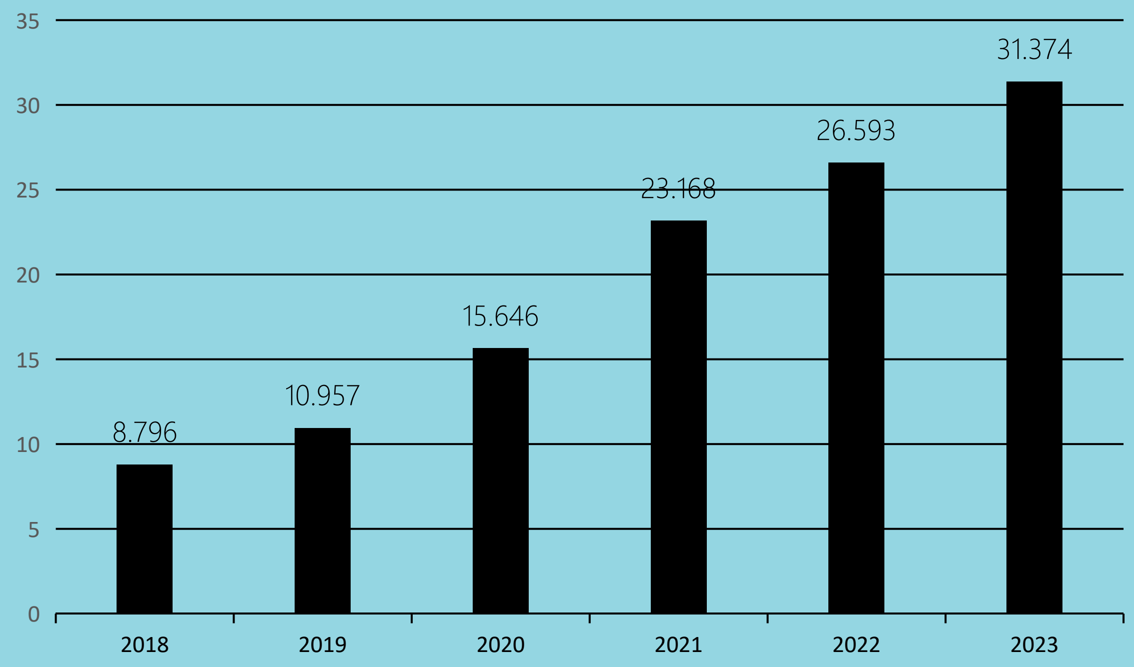
From a quantitative perspective, computing power has a significant impact on the macroeconomic development of a country. According to IDC's report, there is a significant positive correlation between a country's computing power index and its GDP/digital economy trends. On average, for the 15 key countries[1], a 1-point increase in the computing power index leads to a 0.35% growth in the digital economy and a 0.18% growth in GDP. This trend is expected to continue from 2021 to 2025.

*15 key countries: United States, China, Japan, Germany, United Kingdom, France, Canada, South Korea, Australia, India, Italy, Brazil, Russia, South Africa, Malaysia

Computing Power | Strong market demand



Global computing power scale (EFlops)



Global computing power market size (trillion US dollars)

+65%

Annual Growth

The global computing power scale is maintaining a high-speed and stable growth trend. The rapid rise of fields such as artificial intelligence, scientific research, and the metaverse poses higher demands on computing power. It is estimated that by 2030, the average annual growth rate of the global computing power scale will reach 65%.

$31.3B

2023

In 2021, the global computing power network market reached a scale of $23.168 billion, with a year-on-year growth of 48.08%. The computing power network has become a new focus worldwide, and it is expected that the market size will further increase to $31.374 billion by 2023.

Matrix AI

Problem | Restricted AI development

"Computing power has become a key constraint for further development of artificial intelligence"

High computing costs

Ordinary PC is unable to meet the requirements of complex computing tasks. If GPT-3 is used with V100, it would require 355 years of training. Building cloud computing infrastructure requires extremely high costs

4 million USD: The estimated cost of one training session for ChatGPT
0.7 million USD: The estimated daily hardware operating cost for ChatGPT

Inefficient resource utilization

High-end servers are generally scattered, and idle resources are not fully utilized

20-30%: The global computing resource utilization rate

Giant monopoly

Indeed, highly centralized resources and data monopolies give monopolistic companies significant control over data and pricing power for computing power. This can create barriers to entry for smaller players and limit competition in the market

61% Gross margin: AWS offers commoditized compute hardware

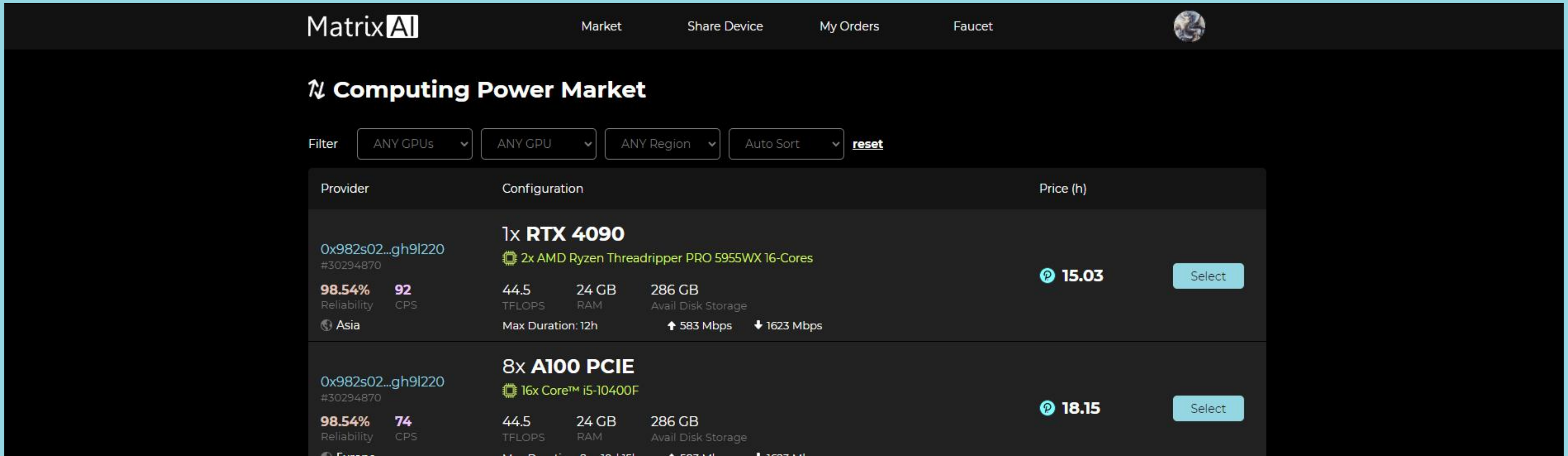
MatrixAI

MatrixAI | Web3 Computing Power Network

MatrixAI includes a **protocol for integrating and scheduling computing power,** as well as a **peer-to-peer trading market**. Its main use case is to allocate computing tasks to idle computing power, and providers receive economic incentives by submitting proof of calculation.

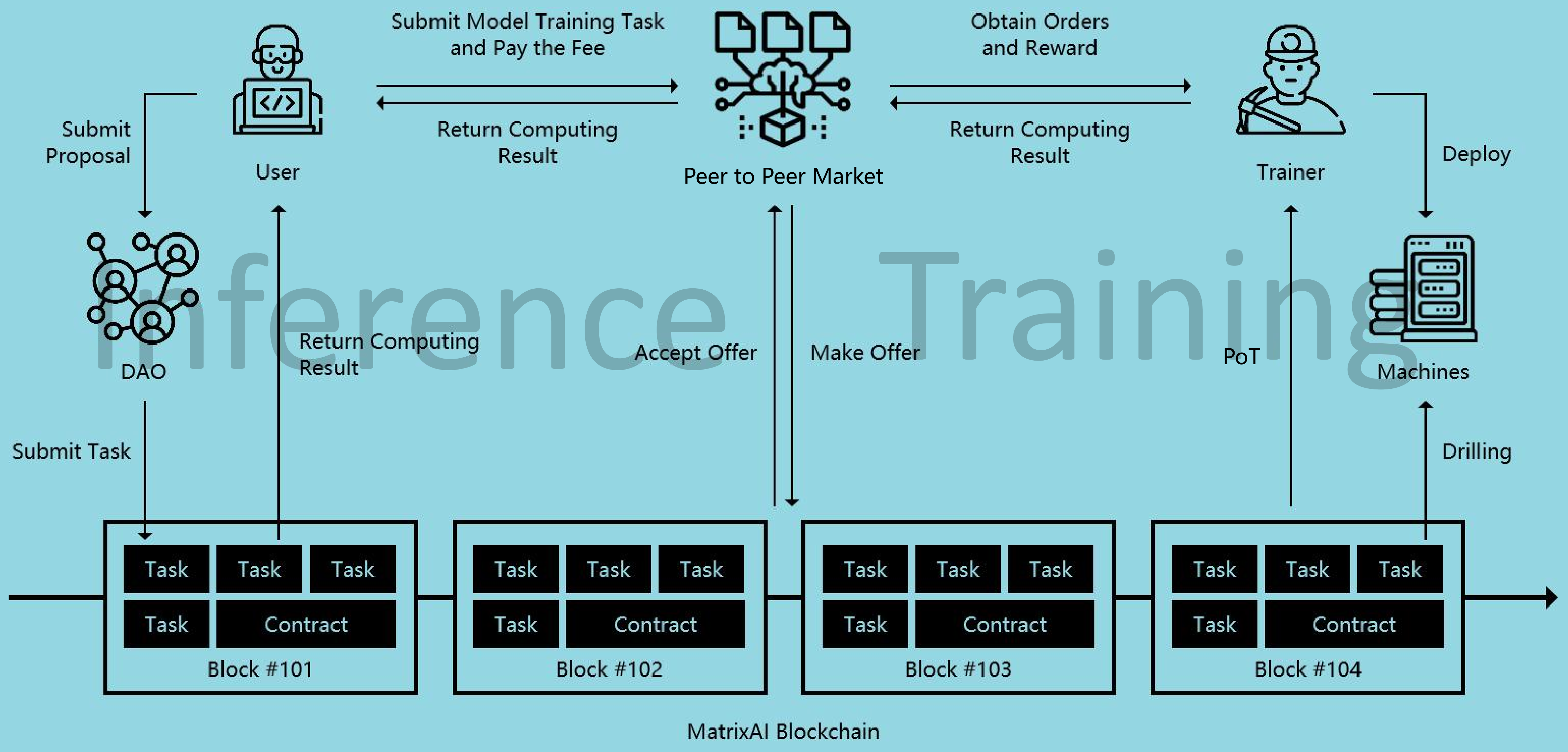
35% **Model Training**

65% **Model Inference**



MatrixAI

Bussiness Model | Public, Transparent and Direct



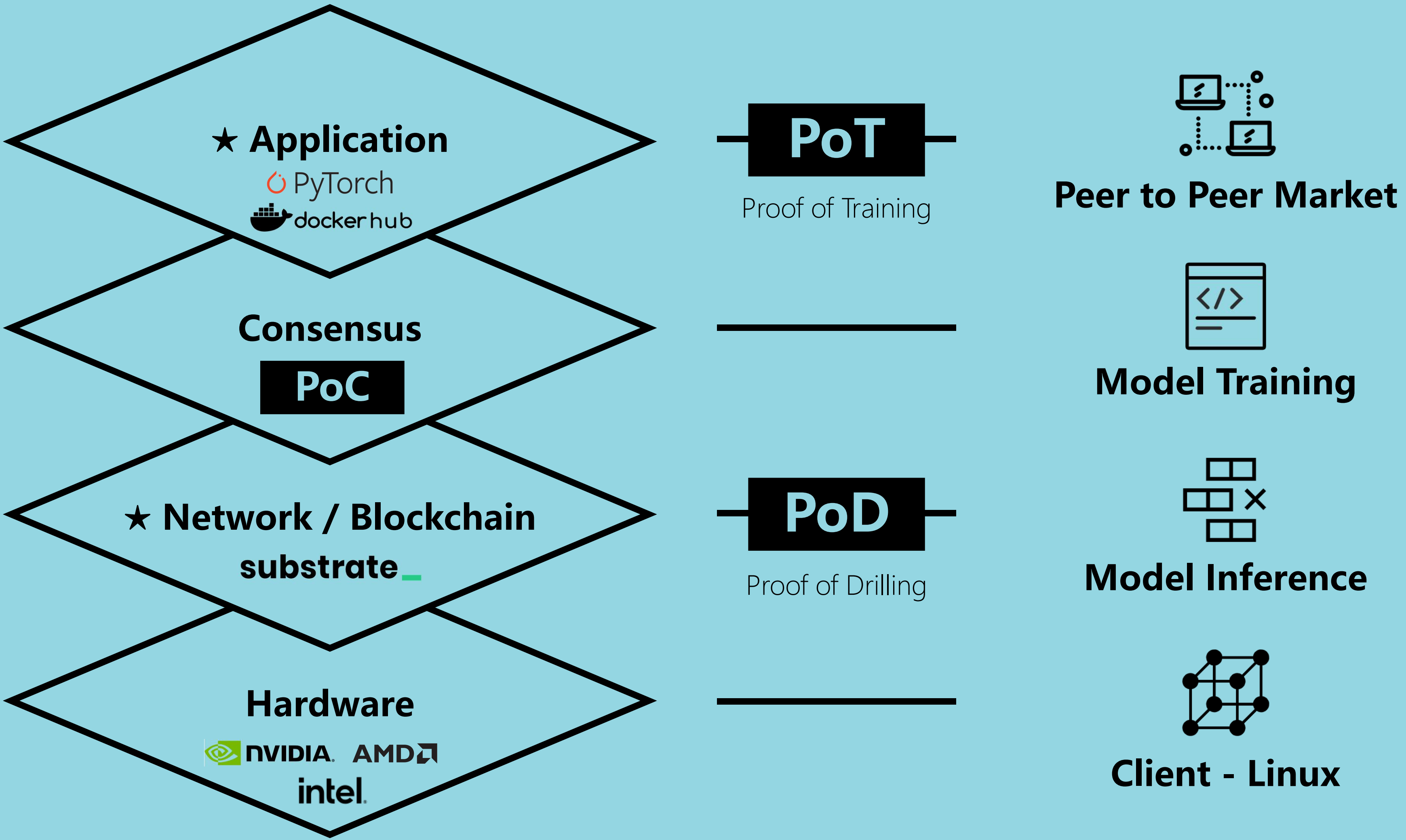
MatrixAI

Computing Network | There is an increasing number of Players

	Feature	Introduction	Investment
exaBITS 2022.4	- Early & Testnet - Apply to access/supply machines - Expected cost reduction by 3x - No economic model - Calculate routing, scheduling, and transactions	exaBITS is a project within Harvard Innovation Lab Incubation Program, aiming to build a Web3.0 infrastructure with zero downtime, low costs, and infinite computing power. The market will have access to various artificial intelligence-related services and resources, including computing services, data storage, and model-as-a-service (MaaS). It can partition large-scale AI models and run them on extremely weak GPUs.	Harvard Innovation Labs
iExec 2016.9	- Mainnet V8 - Connects cloud resource sellers with cloud resource buyers - Computing, Application, data could be traded - Task Orders - TEE	iExec's technology relies on Ethereum smart contracts and enables on-demand provision of high-performance computing services through virtual cloud infrastructure. It allows for on-demand provision of off-chain computing services and datasets, and verification is achieved through the PoCo consensus protocol. iExec also provides other products to enhance ownership and privacy protection, including the Oracle Factory, iExec SDK, and confidential computing.	**$12M**: 2017.4.19 ICO
gensyn 2020.5	- Mainnet - Cloud virtual machine rental - Customizable device parameters - Access to computing resources without permission - Delegated proof of stake consensus mechanism	Cudos is a Layer1 blockchain that focuses on delegated proof of stake. Its vision is to achieve on-chain accessible decentralized computing. The design of its network separates consensus from execution to ensure security, decentralization, and permissionless access to high-performance computing. The design of the Cudos network separates consensus from execution to ensure security, decentralization, and permissionless access to high-performance computing.	**$600K**: 2019.12.4 Seed Round - Outlier Ventures
CUDOS® 2019.6	- Early - Probabilistic Learning Proof - Distributed Parallel Computing - Model Data Privacy Protection - No Economic Model Currently	The Gensyn protocol will become the underlying layer for machine learning computations, similar to the Ethereum platform for executing smart contracts. It aims to be a large-scale and cost-effective computing protocol for global deep learning models. It enables anyone to use a self-organizing network that includes all existing computing resources to train machine learning models for any task.	**Over $50m**: a16z leads the investment, with participation from CoinFund, Canonical Crypto, Protocol Labs, Eden Block

MatrixAI

Difference | Value, Application Scenarios, Layer



Benefit | Efficient, Lower Costs, Fair

More Efficient

- Users can instantly trade with crypto and select a model to start traning in minutes
- Orders are settled almost in real time

Lower Costs

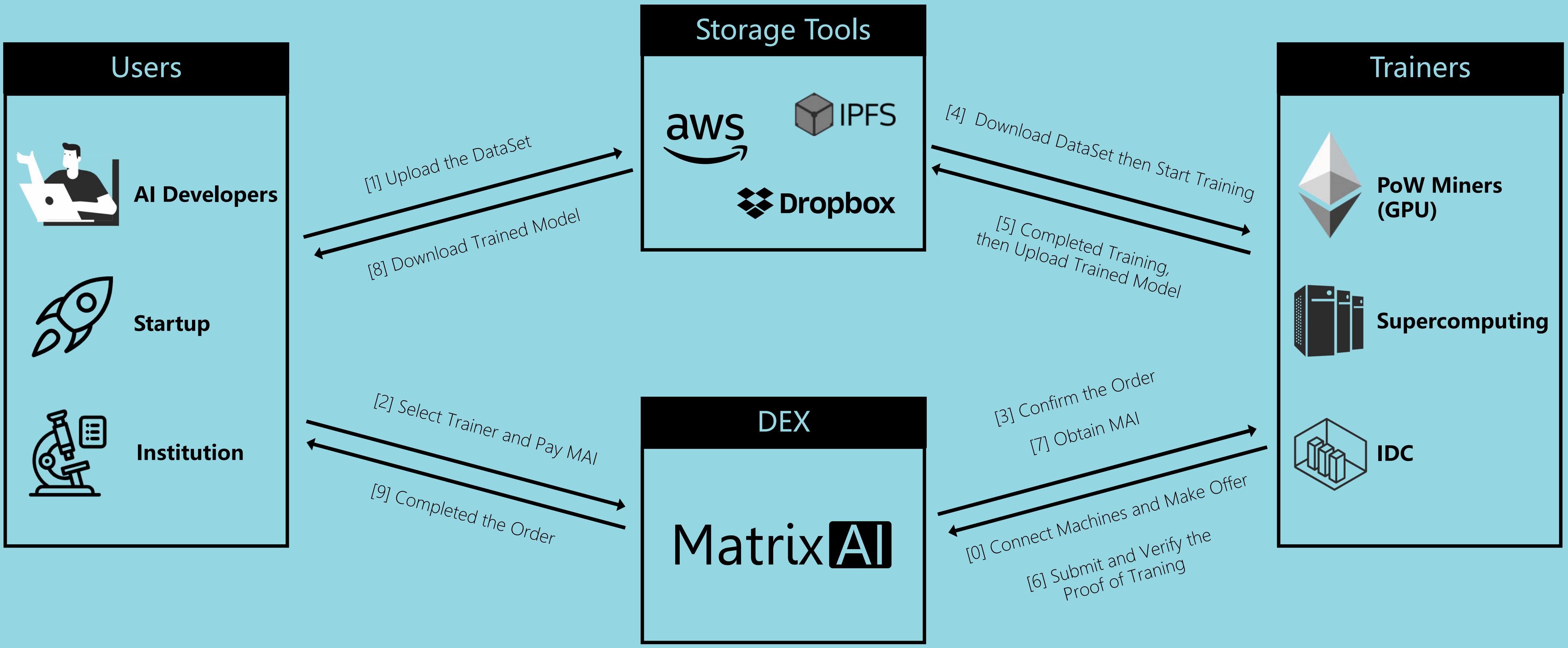
- Introduce bidding from different providers to form price competition. Expected cost is reduction of 75% to 90% or more
- Low access threshold, users can share idle computing power, it means lower costs

The Principle of Fair Incentives

- Incentives are distributed based on the contribution of the participants, rather than distributing profits through resource monopolies

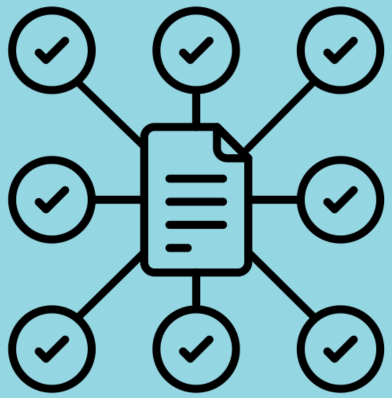
MatrixAI

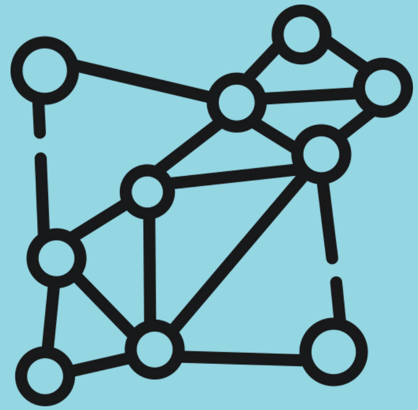
User Case | Connect Users and Trainers



MatrixAI

Solution | Technical Innovation





DEX for Computing Power

By connecting idle computing resources to the MatrixAI Network, trainers can earn economic benefits from their underutilized resources. Users, on the other hand, can send their computational tasks to trainers through an open and transparent marketplace.

Proof of Compute

After completing the computational tasks, trainers can submit the generated PoT or PoD to the blockchain network. The network can randomly assign miners to validate the proof, ensuring that trainers have genuinely and effectively completed the assigned tasks.

Computing integration protocol

The Computing Integration Protocol aims to connect computing facilities with varying geographical locations and configurations to increase the overall computational capacity of the network. The network can leverage the combined power of these facilities, enabling more efficient and scalable computations.

Roadmap | Aggregate Computing Resources

MatrixAI

Q3 2023

- Computer Power (P2P)Market Launched ☑
- Client v1.1 Updating: Support Docker Image ☑
- Completed MatrixAI Litepaper ☑

Q4 2023

- Economical Whitepaper Release
- Mining Machine Testing
- Developer Communication
- Grants Application

Q1 2024

- Computing Integration Protocol
- Client v1.2 Updating: Support More Model

Q2 2024

- Mainnet(Computing Pool) Launched
- MatrixAI Virtual Machine Launched

Litepaper | Upcoming Release

MatrixAI Litepaper

Contract Lab // Released v0.1 // 2023-07

Abstract

The progress of AI heavily relies on large-scale training computation, but the high resource requirements pose barriers to access and impede progress. Current solutions for computational resources are monopolistic, expensive, and impractical for large-scale AI. Decentralized computing services offer advantages such as cost reduction, privacy protection, and innovation opportunities. This paper introduces MatrixAI, a decentralized AI computing marketplace that aggregates idle computing resources worldwide. MatrixAI aims to be the AI computing resource layer network of the Web3 era, supporting diverse scale requirements and breaking centralized monopolies. By empowering computing suppliers and providing cost-effective solutions, MatrixAI shapes the future of decentralized computing and fosters AI innovation.

1 Introduction

1.1 Background

In the past five years, many significant advancements in deep learning have been achieved through the utilization of increasingly large training computation [2,3]. Such large-scale training is accomplished by simultaneously employing hundreds or thousands of specialized accelerators with high internal chip communication bandwidth, such as Google TPUs, NVIDIA A100 and H100 GPUs, or AMD MI250 GPUs. These accelerators are utilized for several weeks or months to compute thousands or millions of gradient updates. Consequently, the tremendous resource requirements involved in constructing these foundational models pose significant barriers to accessing them, and without a means to capture value while sharing resources, it could potentially lead to stagnation in AI progress.

Currently, the solutions available for providing computational resources are either monopolistic and expensive or impractical due to the complex computations required for large-scale AI. Meeting the ever-growing demand necessitates a cost-effective utilization of all available computing resources.

The present challenge lies in the limitations imposed by the asymptotic progress of microprocessor performance on the computational resources themselves, compounded by chip shortages resulting from supply chain and geopolitical factors.

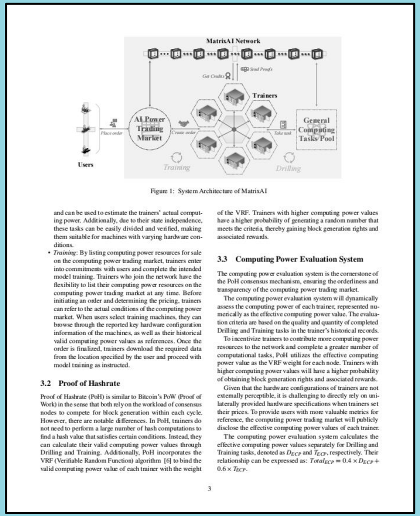
On the other hand, the average utilization rate of global cloud computing data centers has consistently remained low, indicating the presence of a substantial amount of idle computational resources. Furthermore, with the continual rise in computing power of consumer-grade devices, there is untapped potential in the form of idle computing resources from personal computers, servers, and mobile devices used by individuals and businesses. Additionally, decentralized data transmission and storage infrastructures like IPFS, Filecoin, and Storj are constantly improving.

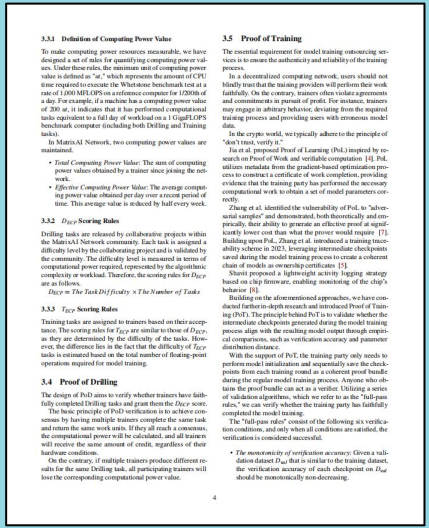
1.2 Motivation

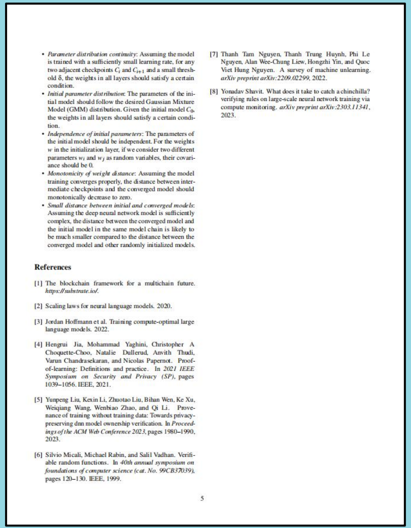
The catalyst for each wave of technological innovation is often the transformation of something expensive into something affordable enough to be wasteful.

Currently, in the field of physical infrastructure, there is a dominant market controlled by vertically integrated giants such as AWS, GCP, Azure, Nvidia, Cloudflare, Akamai, among others. These companies enjoy high-profit margins within the industry. This situation results in high computational costs for new entrants in the AI field, particularly in the LLM domain, which hampers the development and widespread adoption of AI technologies. However, decentralized computing services offer numerous advantages, including decentralized resources, elasticity and scalability, cost reduction, privacy protection, high reliability, as well as opportunities for innovation and collaboration. We firmly believe that decentralized computing services will be the key to overcoming the current high cost of computational power, making it affordable and accessible, thus opening the doors to technological innovation in the AI industry and paving the way for the future of the AI era.

In order to achieve this goal, we have taken the first step







MatrixAI

Core Team | Extensive project experience in AI and WEB3

Dr.Jerry
Technical Scientist

- Holds a Ph.D. degree with a focus on research areas such as information security, blockchain, etc.
- Senior AI expert
- Has extensive experience in large-scale projects related to federated learning and privacy-preserving computation
- Formerly worked at a leading blockchain company

Miles
Technical Expert

- Computer Technology Master, focus on blockchain, distributed storage, and attribute-based encryption
- Has published 3 SCI papers and holds several patents.
- Over 5 years of project management experience
- Worked for a leading blockchain company, leading the research and development of Web3 projects

Han
Product Manager

- EMBAer
- Senior blockchain product expert
- Over 6 years of blockchain product management experience
- Worked in top blockchain company, public chain and crypto currency exchange
- 2022 Polkadot Winter Hackathon 1st Place
- Completed Web3 Foundation Project

Thirteen Chen
Smart Contract Developer

- Has extensive experience in distributed systems and smart contract development
- Previously worked as a senior development engineer at a well-known blockchain company and participated in AI-based smart contract development projects

Matrix AI

Backer | Active groups in the Web3

LD CAPITAL

parity

aws

Google Cloud

Thanks !

We are looking for investors
to accelerate the implementation of AI infrastructure era together

MatrixAI

Contact us
Miles@matrixai.cloud